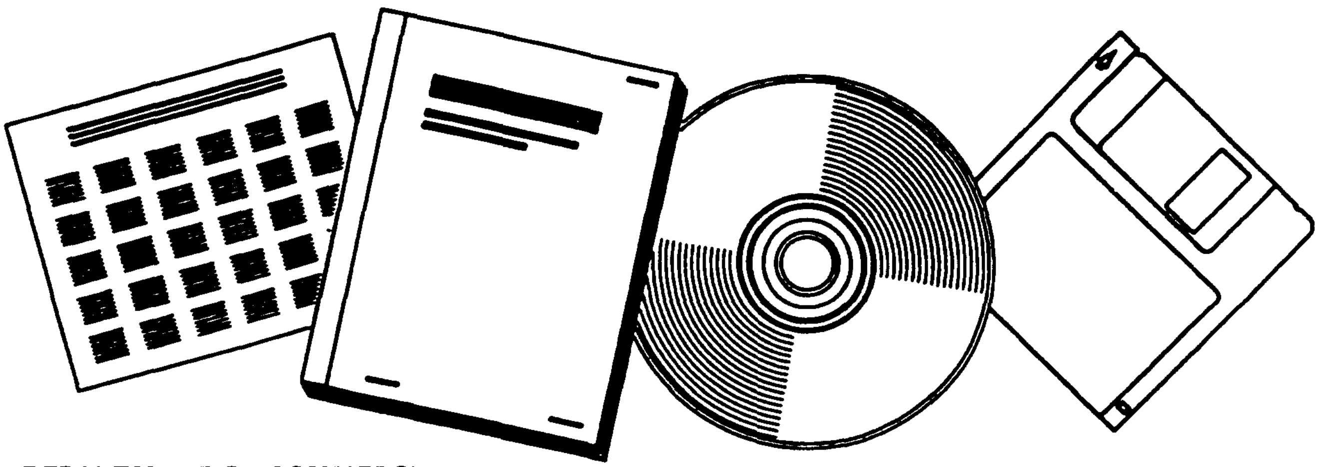
PB97-132286

COMPLEX QUALITY CONTROL OF UPPER-AIR VARIABLES (GEOPOTENTIAL HEIGHT TEMPERATURE, WIND, AND HUMIDITY) AT MANDATORY AND SIGNIFICANT LEVELS FOR THE CARDS DATA SET

(U.S.) NATIONAL CLIMATIC DATA CENTER, ASHEVILLE, NC

21 AUG 96



U.S. DEPARTMENT OF COMMERCE
National Technical Information Service

NATIONAL CLIMATIC DATA CENTER

COMPLEX QUALITY CONTROL OF UPPER-AIR VARIABLES (GEOPOTENTIAL HEIGHT, TEMPERATURE, WIND, AND HUMIDITY) AT MANDATORY AND SIGNIFICANT LEVELS FOR THE CARDS DATA SET

by

Oleg A. Alduchov[1]

and

Robert E. Eskridge

National Climatic Data Center
National Environmental Satellite, Data, and Information Service
National Oceanic and Atmospheric Administration
Asheville, NC 28801-5001

1 Permanent address: Russian Research Institute of Hydrometeorological Information, 6 Korolyov St., Obninsk, Kaluga Reg., 249020, Russia

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
August 21, 1996

3. REPORT TYPE AND DATES COVERED
Final

4. TITLE AND SUBTITLE
Complex quality control of upper-air variables (geopotential height, temperature, wind, and humidity) at mandatory and significant levels for the CARDS dataset

5. FUNDING NUMBERS

6. AUTHOR(S)

Oleg A. Alduchov and
Robert E. Eskridge

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

National Climatic Data Center/NESDIS/NOAA
151 Patton Avenue
Asheville, NC 28801-5001

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Same as above

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Gross (or rough) errors are caused by mistakes or malfunctions at any stage of data processing. Experience suggests that from 5 to 20 percent of upper-air observations contain gross errors. The quality control (QC) procedure's main task is to identify and remove gross errors from the data.

The main problem in developing a reliable QC procedure is developing methods which will minimize both types of quality control procedure errors.

The most complicated and important task in the construction of the Complex Quality Control (CQC) is the development of the DMA. Given the error analysis of each individual CQC component, the DMA must weigh the data in each case and make a decision. The following tests are part of the CQC:

- a comparison of observational data at mandatory levels to horizontal optimal interpolation of data from different stations;
- a comparison of observational data at mandatory levels to vertically interpolated data;
- a check of consistency of mandatory and significant levels for each profile;
- a check that geopotential height and temperature satisfy the hydrostatic equation at mandatory levels;
- a comparison of geostrophic and thermal winds and real winds at mandatory levels;

14. SUBJECT TERMS
quality control   upper-air data
radiosonde

15. NUMBER OF PAGES
154

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

COMPLEX QUALITY CONTROL

OF UPPER-AIR VARIABLES

(GEOPOTENTIAL HEIGHT, TEMPERATURE, WIND, AND HUMIDITY)

AT MANDATORY AND SIGNIFICANT LEVELS

FOR THE CARDS DATA SET

by

Oleg A. Alduchov[1]

and

Robert E. Eskridge

National Climatic Data Center
National Environmental Satellite, Data, and Information Service
National Oceanic and Atmospheric Administration
Asheville, NC 28801-5001

1 Permanent address: Russian Research Institute of Hydrometeorological Information, 6 Korolyov St., Obninsk, Kaluga Reg., 249020, Russia

TABLE OF CONTENTS

Summary

The goal of the "Comprehensive Aerological Reference Data Set" (CARDS) project is to produce an upper-air data set based on radiosonde and pibal observations, suitable for evaluating climate models and detecting global change. The CARDS project is a joint project of the National Climatic Data Center (NCDC) of the United States of America and the All-Union Research Institute of HydroMeteorological Information (AURIHMI), Russia. The CARDS data set has also been identified as a WMO baseline climate data set.

The presence of errors in meteorological data must be taken into account before the data are used. Detecting and removing errors is especially important in any climate change analysis, since the noise (errors) in the observation network and meteorological observations may be larger than the signal (eg temperature change by decade) being investigated.

The three main types of errors in radiosonde data are random observational errors, systematic observational errors, and gross (rough) errors. Observation errors are due to inaccuracies in the measurement of atmospheric variables such as temperature, relative humidity, and pressure. The number and statistical structure of observation errors are determined by the quality of the observations. It is not possible to remove observation errors. But observation errors generally have constant statistical properties and one can take these errors into account by studying their structure.

It is important to differentiate between random and systematic observation errors (Hawson 1970, Hooper 1975). Random and systematic errors are differentiated by their mean value, which is zero for random errors and nonzero for systematic errors. The presence of systematic errors is generally attributed to inadequate or erroneous actions in taking the radiosonde observation, to a change in the instrumentation, or to a change in the data processing procedure. These actions contribute to systematic errors (not necessarily constant in time and space) in the observational data. The detection and removal of these errors is a complicated, but necessary, step in the analysis of climate change.

Gross (or rough) errors are caused by mistakes or malfunctions at any stage of data processing. Experience suggests that from 5 to 20 percent of upper-air observations contain gross errors (Gandin (1988) and Alduchov (1982)). The percentage depends on the part of the world and the time period. The composition, magnitude, and occurrence of particular types of gross errors varies with each dataset. Gross errors can significantly distort the results of any data analysis. Thus, a quality control (QC) procedure is a necessary step in meteorological data processing. The QC procedure's main task is to identify and remove gross errors from the data and it clearly must precede any data analysis.

Systematic errors in time series can be detected by the use of accurate station histories, mathematical-physical models, and/or statistical techniques.

A QC procedure can be logically defined as follows. The variable being controlled is assigned to one of several classes (subsets) into which the set of observations is divided. Usually the data are divided into two classes: a class of correct values, and a class of erroneous values. Errors in the quality control procedure (QCE) occur when the controlled value is assigned to the wrong class. There are two different types of QCEs: a QCE of the 1st type occurs when an erroneous value is assigned to the class of correct values. In a QCE of the 2nd type, a correct value is assigned to the class of erroneous values. It is clear that the occurrence of these errors is highly undesirable, since they may distort any data analysis.

It is not difficult to develop a QC procedure which can minimize the quality control errors of either type. One can apply, for example, a check for physical limits within sufficiently large bounds or gates. This will minimize errors of the 2nd type. The large bounds will guarantee that not a single correct value is taken as erroneous. However, many erroneous values would be taken as correct. To minimize quality control errors of the 1st type it is possible to use the same procedure with very narrow bounds. All erroneous values would be removed, but many correct values would be misclassified as erroneous.

The main problem in developing a reliable QC procedure is developing methods which will minimize both types of quality control procedure errors. Experience suggests that a simple, single-criterion QC procedure cannot minimize the QCE's with the current level of upper-air data redundancy. Thus, more complex methods have to be developed to quality control data from the atmosphere.

The idea of a complex (multicomponent) quality control (CQC) of meteorological data was proposed by L. S. Gandin (1969) and developed under his guidance in other studies: Parfiniewicz (1976), Antsipovich (1980), and Alduchov (1983). This was a new and imaginative approach to solving the problem of meteorological data quality control. Gandin introduced the idea to combine simple quality control methods (CQC components) through a decision making algorithm (DMA) whose working logic would be similar to that of a human being. This integrated system results in increased sensitivity to errors, improved determination of errors, and superior decision making. The CQC minimizes the

number of quality control errors (QCE) of both types without degrading the positive features of each CQC component.

There is little difference between the use of control procedures within the CQC framework and the individual use of each quality control procedure. A criterion, which serves as the basis of a control procedure, is used to check the data. When a suspected value is found, the DMA weighs the analysis of each CQC component, and makes a decision whether the value is correct or erroneous based on a joint analysis of all CQC components. Such a procedure permits the use of significantly smaller bounds.

There are a great variety of errors, and each CQC component has different sensitivities to these errors. Therefore, the most complicated and important task in the construction of the Complex Quality Control (CQC) is the development of the DMA. Given the error analysis of each individual CQC component, the DMA must weigh the data in each case and make a decision.

The choice of quality control components to use in the CQC system is of great importance. For upper-air data, it is useful to check observations for mutual consistency with bracketing soundings (temporal consistency), with adjacent heights (vertical consistency), and with the data of neighboring stations (horizontal consistency). Hence, these types of checks must be components of a CQC for upper-air data. In the context of climate change analysis, the horizontal check is of particular importance since this check will reveal systematic observation errors at individual upper-air stations.

Temporal, vertical, and horizontal checking are usually based on the interpolation of observational data to the station being checked. A comparison is made between the results of the interpolation and the observed values. The data interpolation method plays a significant role in the quality control of upper-air data. There are many mathematical methods used in the interpolation of data. However, optimal interpolation of upper-air data is the preferred method for use in quality control procedures (Gandin, 1963). Optimal interpolation allows not only the accurate interpolation of the data, but gives an estimate of the accuracy of the interpolation at each observation point. Error estimates are used in the quality control procedures. Another advantage of optimal interpolation is that statistics of controlled values (first and second moments) over a field, which are needed for optimal interpolation, are already known from historical data. Therefore, the QC procedures can take into account the historical behavior of the variables being controlled. The more detailed and reliable the statistics used in the interpolation, the more likely is the local behavior of the variable to be controlled correctly.

It is very important during the quality control of upper-air data to make sure a sounding is internally consistent. The main criterion of consistency for geopotential height, temperature, and pressure is the requirement that the hydrostatic equation be satisfied. The hydrostatic equation is the basis for one of the most effective QC methods for upper-air data.

Tests for internal consistency of geopotential heights, temperatures, and winds are provided by checking the data against the geostrophic and thermal wind equations. These tests use optimal differentiation of the geopotential and temperature fields (Gandin and Kagan, 1976).

The ability of quality control to detect and to locate errors in the data depends on the skill to create an accurate prediction of the value in question, and the skill to use several independent predictions of the value. The more accurately one can calculate (predict) the value in question, the smaller the errors that can be detected. If there is only a single predicted value of an observation, one cannot be sure which is erroneous: the observation, the prediction, or perhaps both. As a rule, to calculate a predicted value for an observation, observations which are questionable must be used. Therefore, it is necessary to have several independent predictions of each observation to accurately locate erroneous observations.

To quality control the CARDS' upper-air data, a complex quality control (CQC) method has been developed which allows us to check geopotential height, temperature, wind speed and direction, and humidity at mandatory and significant levels. The following tests are part of the CQC:

- a comparison of observational data at mandatory levels to horizontal optimal interpolation of data from different stations;
- a comparison of observational data at mandatory levels to vertically interpolated data;
- a check of consistency of mandatory and significant levels for each profile;
- a check that geopotential height and temperature satisfy the hydrostatic equation at mandatory levels;
- a comparison of geostrophic winds and real winds at mandatory levels;
- a comparison of the thermal wind to the real wind at mandatory levels.

I. Main principles of upper-air data quality control

1. Errors in upper-air data.

Before upper-air data are used, errors in meteorological data must be accounted for. Removing and detecting errors is especially important in any climate change analysis, since the noise (errors in the meteorological observations) may be larger than the quantity (e.g., temperature change by decade) being investigated.

The three main types of errors in radiosonde data are random observational errors, systematic observational errors, and gross (rough) errors. Observational errors are due to inaccuracies in the measurement of atmospheric variables such as temperature, relative humidity, and pressure. The number and statistical structure of observational errors are determined by the quality of the observations. It is not possible to remove random observational errors. Random observation errors have relatively constant statistical properties and one can take these errors into account by studying their structure.

It is important to differentiate between random and systematic observational errors (Hawson 1970, Hooper 1975). Random and systematic errors are differentiated by their mean value, which is zero for random errors and nonzero for systematic errors. The presence of systematic errors is generally attributed to inadequate or erroneous actions in taking the radiosonde observation, to a change in the instrumentation, or a change in the data processing procedure. These actions cause the emergence of systematic errors (not necessarily constant in time and space) in the observational data. The detection and removal of systematic errors is a complicated, but necessary, step in the analysis of climate change.

Gross (or rough) errors are caused by mistakes or malfunctions at any stage of data processing. Experience suggest that from 5 to 20 percent of upper-air observations contain gross errors (Gandin (1988) and Alduchov (1982)). The percentage depends on the part of the world and the time period. The composition, magnitude, and occurrence of particular types of gross errors varies with each dataset. Gross errors can significantly distort the results of any data analysis. Thus a quality control (QC) procedure is a necessary step in meteorological data processing. The QC procedure's main task is to identify and remove gross errors from the data and it clearly must precede any analysis.

2. Methods of quality control checks.

A QC procedure can be logically defined as follows. The variable being checked is assigned to one of several classes (subsets) into which the set of observations is divided. Usually the data are divided into two classes: a class of correct values and a class of erroneous values. Errors in the quality control procedure (QCE) occur when the value is assigned to the wrong class. There are two different types of QCEs: a QCE of the first type occurs when an erroneous value is assigned to the class of correct values. In a QCE of the second type, a correct value is assigned to the class of erroneous values. It is clear that the occurrence of these errors is highly undesirable, since they may distort any data analysis.

It is not difficult to develop a QC procedure which can minimize the quality control errors of either type. One can apply, for example, a check for physical limits within sufficiently large bounds or gates. This will minimize errors of the second type. The large bounds will guarantee that not a single correct value is taken as erroneous. However, many erroneous values would be taken as correct. To minimize quality control errors of the first type, it is possible to use the same procedure with very narrow bounds. All erroneous values would be removed, but many correct values would be misclassified as erroneous.

The main problem in developing a reliable QC procedure is developing methods which will minimize both types of quality control procedure errors. Experience suggests that a simple single criterion QC procedure can not minimize the QCEs with the current level of upper-air data redundancy. Thus, more complex methods have to be developed to quality control data from the atmosphere.

QC methods for upper-air data are based on some redundancy in the data.

Redundancy in the data is given by natural laws which define the space-time distribution of upper-air variables. The presence of statistical, dynamic, and thermodynamic laws leads to the development of spacial and time correlations for each thermodynamic variable and relationships between these variables. The more relationships that can be developed, the better the data can be quality controlled.

The simplest and most widely used method of quality control check of upper-air data is a check of allowable values. This method is based on climatological information. For example, a temperature dataset can be tested based on our knowledge that temperature can range from -80 to 50 °C in the lower atmosphere. But this type of QC can detect only the largest gross errors. A natural extension to this QC method is to extend the check to different heights in the atmosphere, seasons of the year, and different locations.

The next step in complexity is to use knowledge of the statistics of each variable to develop quality control checks. The observed value is compared to the mean and the standard deviation. This type of QC will have large error bounds and therefore is usually used as a rudimentary check. It is often used as the first check in a quality control program.

More advanced QC methods that have a narrow "gate" for gross errors are based on the continuity of upper-air variables, e.g. neighboring values should be "close" to the value being tested or they are based on relationships such as the hydrostatic equation. Advanced QC methods include horizontal, vertical and temporal checks of the data. In these advanced methods, the value of the observation is estimated by interpolating from adjacent levels (vertical check), from neighboring upper-air stations (horizontal check), or from consecutive soundings (temporal check). The interpolated value, f_i, is compared to the observed value, f_o. If the absolute value of the difference δf, called the "actual" residual and defined in (1), is small

$$\delta f = | f_0 - f_i | \tag{1}$$

then the observed value is considered to be correct. If δf is large, we assume that the observed value, f_o, is erroneous. To use this test, a criterion must be developed to establish acceptable levels of δf.

The data interpolation method plays a significant role in the quality control of upper-air data. Optimal interpolation of upper-air data is the preferred method to use in quality control procedures (Gandin, 1963). Optimal interpolation produces not only an accurate interpolation of the data (the best in a root-mean-square (rms) sense for normally distributed variables), but gives an estimate of the accuracy of the interpolation at each observation point. These error estimates are used in the quality control procedures to define an acceptable level of differences, δf, between the interpolated and observed values. Optimal interpolation has another important advantage compared to other interpolation methods. The advantage of optimal interpolation is that the statistics (first and second moments) of each variable, which are needed for optimal interpolation, are already known from historical data. Therefore, the QC procedures can take into account the historical behavior of the variables being checked. The more detailed and reliable the statistics used in the interpolation, the more likely is the local "behavior" of the variable to be treated correctly.

There are many other methods used to interpolate and extrapolate data, for example, spline interpolation, polynomial interpolation, etc. However, interpolations with these methods can give poor results, because they are usually based on some artificial rules of data distribution which are not true for all atmospheric states. Whereas optimal interpolation, which uses our knowledge of the atmosphere, very seldom produces inaccurate interpolations of the data.

Using incomplete or an incorrect statistical structure of the atmosphere will limit optimal interpolation's accuracy. Optimal interpolation methods do not react to all possible types of errors and tend to spread the distribution of the actual residuals, because the actual residuals may be large due to some erroneous data being used in the interpolation. Increasing the spread of the actual residuals makes it more difficult to decide if the value is erroneous or correct.

A very important group of QC methods for upper-air data uses the laws of atmosphere physics to test one or more upper-air observations. For example, one of most effective QC methods for geopotential height and temperature is based on the hydrostatic equation. To check wind data, QC methods based on comparing the real wind with the geostrophic or thermal wind can be used.

Each QC method mentioned above is based upon using only a single correlation or equation. Each method reacts to only certain error types, or it has a low sensitivity to gross errors, or does not accurately locate erroneous values. Therefore, upper-air data processing centers use different QC methods sequentially at different stages of the data processing.

3. Complex quality control check of upper-air data.

The concept of a complex (multicomponent) quality control (CQC) check of meteorological data was proposed by Gandin (1969) and developed under his guidance in other studies: Parfiniewicz (1976), Antsipovich (1980), and Alduchov (1983). This was a new and imaginative approach to solving the problem of meteorological data quality control check. Gandin introduced the idea to combine, through a decision making algorithm (DMA), simple quality control methods (CQC components) into a complex system, whose working logic would be similar to those of a human. This integrated system results in increased sensitivity to errors, improved determination of errors, and superior decision making. The CQC minimizes the number of quality control errors (QCE) of both types without degrading the positive features of each CQC component.

There is little difference between the use of procedures within the CQC framework and individual use of each quality control procedure. A criterion, which serves as the basis of a quality control procedure, is used to check the data. When a suspicious value is found, the DMA weighs the analysis of each CQC component and makes a decision whether the value is correct or erroneous based on a joint analysis of all CQC components. Such a procedure permits the use of significantly smaller bounds for each CQC component.

There is a great variety of gross errors and each CQC component has different sensitivities to these errors. Therefore, the most complicated and important task in the construction of the Complex Quality Control check (CQC) is the development of the DMA. From the error analysis of each individual CQC component, the DMA must weigh the data in each case and make a decision.

Upper-air data can be organized into time series or synoptic sort. The organization of the data must be considered as CQC development requires choosing the appropriate CQC components. Upper-air data in synoptic sort is for one hour and usually the whole world. With data in synoptic sort, it is possible to use various horizontal checking techniques and it is nearly impossible to use temporal methods. With time series the inverse problem occurs, it is possible to apply various temporal QC methods and impossible to apply horizontal checks. It is possible to apply the various vertical checking methods to data in either synoptic or temporal sort.

The hydrostatic equation is the most powerful relation to use in the quality control check of geopotential height and temperature. It may be used on data in either synoptic or temporal sort.

Horizontal interpolation, geostrophic and thermal wind relation can be applied to winds in synoptic sort. Wind data in time sort can only be interpolated.

The statistical structure of the variables in the atmosphere must be known to develop QC methods. The calculation of the structure is a complex and time consuming task. An additional difficulty in developing these QC methods is that the statistical structure is usually given only at mandatory levels. It should be noted that in the development of quality control methods, progress has been mainly due to the requirements of weather forecasting, which needs high quality data at mandatory levels. This explains why the quality control methods of data at mandatory levels have advanced the furthest.

A natural step is to develop QC methods which check mandatory level data with significant level data. The use of this type of QC is limited by the fact that data at significant levels can contain errors and these data need to be quality controlled. For this reason, the most advanced methods of quality control check of data at mandatory levels are based on using data from other mandatory levels and not from significant levels. After the mandatory level data has been quality controlled using mandatory data, these data will be used to quality control the significant level data. However, experience shows using data from significant levels is often very helpful for the quality control of data at mandatory levels for complex situations such as at the tropopause or in the boundary layer.

There is data redundancy in mandatory level observations which have a typical horizontal scale of 300-500 km, vertical scale of 1.5-3.0 km, and a time scale of 6-12 hours. Using this data redundancy, the CQC is able to detect temperature errors of 5 to 10 °C, geopotential height errors of 30-60 gpm, wind speed errors of 10-15 m/s, and dewpoint depression errors of 10-15 °C. Since the quality controlled mandatory data has the above accuracy, it is not possible to produce a radiosonde profile with greater accuracy. The accuracy of a profile is defined by its least accurate element. Therefore, the accuracy of quality controlled significant level data is defined by the accuracy of mandatory level data. To quality control data at significant levels with the accuracy of mandatory levels, it is sufficient to use simple methods of interpolation such as linear interpolation. The use of simple methods is possible because the interpolation distance between significant levels and mandatory levels is approximately half the interpolation distance between mandatory levels.

To quality control upper-air observations in synoptic sorted files, a complex quality control method composed of the following components is recommended:

- a hydrostatic check of temperature (T) and geopotential height (H) at mandatory levels;
- a horizontal check of T at mandatory levels based on optimal interpolation of T from neighboring stations;
- a horizontal check of H at mandatory levels based on optimal interpolation of H from neighboring stations;
- a horizontal check of geopotential height at mandatory levels based on optimal interpolation of thickness between mandatory levels from neighboring stations;
- a vertical check of T at mandatory levels based on optimal interpolation of T from adjacent mandatory levels;

- a vertical check of H at mandatory levels based on optimal interpolation of H from adjacent mandatory levels;
- a horizontal check of wind components at mandatory levels based on optimal interpolation of the wind components from neighboring stations;
- a vertical check of wind components at mandatory levels based on optimal interpolation of data from adjacent levels;
- a geostrophic check of wind components at mandatory levels based on optimal differentiation of H from neighboring stations;
- a thermal wind check of wind components at mandatory levels based on optimal differentiation of T from neighboring stations and adjacent levels;
- a horizontal check of humidity, R, at mandatory levels based on optimal interpolation of R from neighboring stations;
- a vertical check of humidity, R, at mandatory levels based on optimal interpolation of R from adjacent mandatory levels;
- a vertical check of T at significant levels based on linear interpolation of T from adjacent mandatory levels;
- a vertical check of H at significant levels, based on linear interpolation of H from adjacent mandatory levels;
- a vertical check of wind at significant levels based on linear interpolation of wind from adjacent mandatory levels;
- a vertical check of R at significant levels based on linear interpolation of R from neighboring mandatory levels;

To quality control upper-air observations in station sort (time series), a complex quality control method based on the following components is recommended:

- a hydrostatic check of T and H at mandatory levels;
- a temporal check of T at mandatory levels based on optimal interpolation of T from consecutive observation hours;
- a temporal check of H at mandatory levels based on optimal interpolation of H from consecutive observation hours;

- a temporal check of geopotential thickness at mandatory levels based on optimal interpolation of thickness between mandatory levels from neighboring stations;
- a vertical check of T at mandatory levels based on optimal interpolation of T from adjacent mandatory levels;
- a vertical check of H at mandatory levels based on optimum interpolation of H from adjacent mandatory levels;
- a temporal check of wind components at mandatory levels based on optimum interpolation of wind components from consecutive observation hours;
- a vertical check of wind components at mandatory levels based on optimal interpolation of data from adjacent levels;
- a temporal check of R at mandatory levels, based on optimal interpolation of R from consecutive observation hours;
- a vertical check of R at mandatory levels based on optimal interpolation of R from adjacent mandatory levels;
- a vertical check of T at significant levels based on linear interpolation of T from adjacent mandatory levels;
- a vertical check of H at significant levels based on linear interpolation of H from adjacent mandatory levels;
- a vertical check of wind at significant levels based on linear interpolation of wind from adjacent mandatory levels;
- a vertical check of R at significant levels based on linear interpolation of R from adjacent mandatory levels.

The two proposed schemes for the Complex Quality Control check of synoptic and station sorted data will enable us to detect many types of gross errors in the upper-air data. The exclusion of any component of the CQC will degrade our ability to detect these errors.

The hydrostatic quality control check, as previously noted, is a very powerful method for the quality control of temperature and geopotential height data. It allows a simultaneous check of temperature and geopotential height and the hydrostatic equation can be used to calculate a correct temperature and geopotential height once an error has been detected. However, a hydrostatic quality control check does not respond to certain types of errors in the data. It can not locate erroneous data at the top or bottom of a radiosonde sounding, and it will not detect some combinations of errors. Experience shows that a hydrostatic quality control check will locate and correct approximately 55% of the errors in the temperature and geopotential height data at mandatory levels.

A hydrostatic quality control check, together with a horizontal check for synoptic sorted data or a temporal check for station sorted data, remarkably improves our ability to detect and locate errors. A horizontal check gives the greatest improvement in a dense network of upper-air stations. A temporal check gives the greatest improvement at stations taking frequent observations. For a sparse dataset, horizontal and temporal checks have approximately the same skill as climatic check and they can detect only very large errors. Sometimes horizontal and temporal check methods are based on an interpolation using geopotential height rather than geopotential thickness. Changing the interpolation from geopotential height to thickness improves the detection of errors to such degree that horizontal and temporal checks are then almost as sensitive as the hydrostatic quality control check. But horizontal and temporal interpolation of geopotential height can detect certain errors that thickness can not detect, e.g. distortion of all heights by a constant value. Hence, all these methods should be part of a quality control procedure.

A vertical quality control check is quite beneficial when the data are sparse in space or time. Due to the constant distribution of mandatory levels, the skill of a vertical check in detecting errors at mandatory levels depends mainly on the variable and its behavior (statistics). The sensitivity of a vertical control to errors in geopotential thickness and temperature data are close to that of the hydrostatic quality control check. This is true even for geopotential heights, because of the high correlations between adjacent mandatory levels.

For difficult situations such as the tropopause, the boundary layer, and where local conditions strongly affect the data, it is very beneficial to use significant level temperature data.

The CQC of geopotential thickness and temperature data by a combination of hydrostatic, horizontal, temporal, and vertical checks allows the CQC to detect errors in both dense and sparse sets of radiosonde stations. The use of horizontal and temporal interpolation of geopotential thickness is highly sensitivity when used in combination with other checking components.

The CQC scheme for wind data uses horizontal, vertical, or temporal quality checks. For the wind, there does not exist a powerful and accurate relation like the hydrostatic equation. Hence, the ability to detect errors in wind data is relatively low. For a sparse set of upper-air stations, the quality control checks are a vertical and climatic check. For a dense set of synoptic sorted data, error detection can be improved using geostrophic and thermal wind checks. The geostrophic and thermal wind checks give a small improvement in the quality control of winds. More important, these relationships give another check on the consistency of geopotential height, temperature, and the wind data. Wind data are also checked using significant level data.

The methods for controlling humidity data are limited. A combination of horizontal, vertical, or temporal checks can be used. The skill of these methods is limited, because of the high variability of humidity in atmosphere and the many large observational errors in humidity at low temperatures. Some small improvement is achieved by adding a pure statistical component, which is based on the correlations between humidity and temperature.

It is important to use the data from significant levels to quality control humidity. First, there are few other choices; second, humidity is highly variable in the vertical direction.

In summary, the CQC procedure checks the data at mandatory levels, then performs a vertical quality check of temperature, humidity, and wind at significant levels using the checked data from mandatory levels.

Examples of the response of the CQC components to different simple gross errors in the data at mandatory levels are shown in Table 1. This table maybe expanded to include the complete relationship between the actual residuals of the CQC components to the location and value of an error. This will then form a good base for building a decision making algorithm (DMA). The DMA solves the inverse task of predicting the correct value given the location and size of an error using the actual residuals from the CQC components.

II. The components of the complex quality control of upper-air data

1. Hydrostatic quality control check

Integration of the hydrostatic equation

$$\frac{\partial H}{\partial P} = -\frac{R\ T_V}{g\ P}, \qquad (2)$$

where H is the geopotential height, P the is pressure, R is the gas constant for dry air, g is the acceleration due to gravity, T_V is the virtual temperature for the layer between two mandatory levels with pressure P_i and P_{i+1}, and assuming that virtual temperature and the acceleration due to gravity are constants in this layer, yields the hypsometric equation

$$H_{i+1} - H_i = \frac{R}{g}\ T_{V_i}^{i+1} \ln\frac{P_i}{P_{i+1}}, \qquad (3)$$

where T_{Vi}^{i+1} represents the averaged virtual temperature for the P_i-P_{i+1} layer.

The hypsometric relation is used at the radiosonde station to calculate geopotential height, H_i, at mandatory levels when the sounding is produced. The averaged virtual temperature for each layer is calculated using the temperature and humidity data from mandatory and significant levels for this layer.

Replacing T_{Vi}^{i+1} in (3) by

$$T_i^{i+1} = (T_i + T_{i+1})/2, \qquad (4)$$

where T_i and T_{i+1} are the temperatures at the lower and upper mandatory levels, then the hydrostatic residual form is defined as

$$\delta_i^{i+1} = H_{i+1} - H_i - B_i^{i+1}(T_i + T_{i+1}), \tag{5}$$

where

$$B_i^{i+1} = \frac{R}{2g} \ln \frac{P_i}{P_{i+1}}. \tag{6}$$

The residual in (5) represents the error made in using temperature instead of virtual temperature together with the reported geopotential heights H_i and H_{i+1}.

Research shows that mean values of the δ_i^{i+1} are small compared to most gross errors for the layers between mandatory levels. The rms values defined by

$$E_i^{i+1} = \sqrt{\overline{(\delta_i^{i+1})^2}} \tag{7}$$

are small and depend weakly on season and latitude.

Table 2 contains mean and rms values of the hydrostatic residual (5). They have been calculated for different latitude zones and different layers between 1000 and 10 hPa using data from 0 UTC, 15 January 1989. The residuals in Table 2 are expressed in geopotential meters. Table 3 shows the normalized mean and rms values attributed to temperature (that is $\alpha=\delta_i^{i+1}/B_i^{i+1}$ and $\beta = E_i^{i+1}/B_i^{i+1}$). Table 3 shows the difference between averaged virtual temperatures and temperatures calculated by (4), because (5) ignores humidity and temperature at significant levels.

Fig. 1 shows the distribution of the normalized hydrostatic residuals ($\delta_i^{i+1}/\Delta_i^{i+1}$) for the 850-700, 500-400, and 100-50 hPa layers for a global set of upper-air data calculated by the hydrostatic check. The data in Fig. 1 shows that correct values of T_i, T_{i+1}, H_i and H_{i+1} will satisfy

$$|\delta_i^{i+1}| \le \Delta_i^{i+1} = 4.0 \times E_i^{i+1} \tag{8}$$

almost 100% of the time. This means that, if inequality (8) is not satisfied, at least one of values T_i, T_{i+1}, H_i or H_{i+1} is erroneous.

If H_i contains error χ, then for two adjacent layers the following approximations can be found from (5)

$$\delta_{i-1}^{i} \sim -\delta_{i}^{i+1} \sim \chi \tag{9}$$

Hence, the error χ can be approximated for interior levels by

$$\chi = \frac{\delta_{i-1}^{i} - \delta_{i}^{i+1}}{2} \tag{10}$$

and as

$$\chi = \delta_{n-1}^{n} \tag{11}$$

and

$$\chi = -\delta_{2}^{1}, \tag{12}$$

for the top and bottom levels.

For error τ in T_i, it follows from (5) for two adjacent layers that

$$\delta_{i-1}^{i}/B_{i-1}^{i} \sim \delta_{i}^{i+1}/B_{i}^{i+1} \sim \tau \tag{13}$$

Hence, we can approximate the error τ in T_i for interior levels as

$$\tau = \frac{\delta_{i-1}^{i}/B_{i-1}^{i}+\delta_{i}^{i+1}/B_{i}^{i+1}}{2}, \tag{14}$$

and

$$\tau = \delta_{n-1}^{n}/B_{n-1}^{n} \tag{15}$$

at the top level, and

$$\tau = \delta_2^1 / B_2^1 . \quad (16)$$

for the bottom level.

Fig. 2 shows the rms values of differences between real errors and calculated errors using equations (10) through (12) and equations (14) through (16). Fig. 2 shows that the most accurate estimates of the errors occur at intermediate levels with larger error estimates at the surface and the higher atmospheric levels. Fig. 2 shows there are systematic errors in the residual mean values in temperature in the lower levels. This is due to the use of temperature instead of virtual temperature in the hydrostatic equation. A simple way to improve the results of the hydrostatic check of geopotential height and temperature data is to use the humidity data.

2. Horizontal and vertical interpolation

Upper-air thermodynamic variables are continuous in time and space. Continuity ensures that the difference between two observations taken at nearby points is small. Quality control procedures can be developed based on these facts. Quality control checks based on continuity consists of comparing observations at mandatory levels for a given station at a given time, with interpolated values from neighboring stations (horizontal check); interpolation from adjacent levels at the same time at the same station (vertical check); or interpolation at a level using consecutive observations in time from the same station (temporal check).

As discussed earlier, optimal interpolation is the best method to use in the quality control of upper-air data when the first and second moments are accurately known. An advantage of the optimal interpolation method is the ability to calculate not only differences between observed and interpolated values (residuals), but also allowable value ranges.

2.1 Optimal interpolation assuming no observational errors

Let f'_i be the departure from the monthly mean, $\bar{f}_i$, where f_i represents the observations in four-dimensional time and space i (i = 0, 1, ..., n), and σ_i is the standard deviation of the observations.

We will interpolate or extrapolate f'_i (i=1,...,n) to a point 0 using

$$\hat{f}'_0 = \sum a_i \frac{\sigma_0}{\sigma_i} f'_i. \tag{17}$$

The coefficients a_i (i=1,...n) are defined so that the square of the rms of the difference,

$$\delta_0 = f'_0 - \hat{f}'_0, \tag{18}$$

is a minimum at each point, that is,

$$E^2 = \overline{\delta_0^2}, \tag{19}$$

is to be minimized.

Necessary conditions for a minimum are

$$\frac{\partial E^2}{\partial a_i} = 0 \ , \qquad i = 1, \ldots, n. \tag{20}$$

Using (17) through (19), E^2 is found to have the following form

$$\begin{aligned} E^2 &= \overline{(f'_0 - \sum_{i=1}^{n} a_i \frac{\sigma_0}{\sigma_i})^2} \\ &= (\overline{f'_{0_2}} - 2\sum_{i=1}^{n} a_i \frac{\sigma_0}{\sigma_i} \overline{f'_0 f'_i} + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \frac{\sigma_0{}^2}{\sigma_i \sigma_j} \overline{f'_i f'_j}) \\ &= (1 - 2\sum_{i=1}^{n} a_i \mu_{0i} + \sum_{i=1}^{n}\sum_{j=1}^{n} a_i a_j \mu_{ij}) \ \sigma_0^2, \end{aligned} \tag{21}$$

where

$$\mu_{ij} = \frac{\overline{f_i^\prime f_j^\prime}}{\sigma_i \sigma_j} \tag{22}$$

is the correlation of the observations, f, at the i and j points.

Equation (20) together with (21) yields a system of linear equations

$$\sum_{i=1}^{n} a_i \mu_{ij} = \mu_{0j}, \qquad j=1,\ldots,n, \tag{23}$$

which are solved for a_i. Now (17) can be used to produce a solution (interpolated value) which is a minimum, in an rms sense, of the difference between the observed f' and interpolated $\hat{f}'$ values.

With optimal interpolation the rms values of the difference between observed and interpolated values can be calculated. Equations (21) and (23) yield

$$E = \sigma_0 \ (1 - \sum_{i=1}^{n} a_i \mu_{0i})^{0.5}. \tag{24}$$

If it is assumed that the distribution is normal, the allowable difference or residual, Δ_0, can be estimated using interpolated and observed values as

$$\begin{aligned} \Delta_0 &= N_f \ E \\ &= N_f \ \sigma_0 \ (1 - \sum_{i=1}^{n} a_i \mu_{0i})^{0.5}, \end{aligned} \tag{25}$$

where N_f is a constant, assumed to be equal to 4.0, this value can be modified after the data are analyzed.

The method of optimal interpolation is similar to the method of linear multiple regression. The main difference between them is that in multiple regression all of the necessary statistical characteristics are calculated using the data, whereas in the optimal interpolation method statistical properties of each parameter must be known beforehand.

2.2 Horizontal optimal interpolation of upper-air variables

One of the principal difficulties in using horizontal optimal interpolation is finding a set of stations whose data is highly correlated with the data of the station being tested. Theoretically, one can take all stations, but the number of radiosonde stations is very large (hundreds). In this case, solving the resulting system of linear equations to determine the weights, a_i, will be difficult. Moreover, the contribution of most of the stations to the calculated weights is negligible. Gandin and Kagan (1976) and Liberman (1980) have shown that using four to eight neighboring stations in horizontal optimal interpolation of upper-air variables yields the most accurate results. An even distribution around the interpolation point is ideal.

In this quality control scheme, eight stations will be used in the interpolation. For interpolating stations, only those which are located within 2000 km of the given station are considered. To find an even and symmetrical station distribution, a circle of influence (2000 km radius) with eight 45 degrees sectors is constructed. From each sector, no more than two stations are picked. Any of these 16, or less, stations can be used in the optimal interpolation.

The eight "influencing" stations are chosen by the following procedure:

a) In each of the eight sectors, the station whose data has the greatest correlation with data from the given station is picked;

b) if any sectors are empty, then a station is chosen from a sector which contains more than one station;

c) if the number of selected stations in the eight sectors is less than or equal to eight, then all stations are used in the optimal interpolation;

d) weights, actual residuals, and allowable residuals are calculated at all levels;

e) if the station being tested has data which is lacking in one or more of the influencing stations, then a new set of eight influencing stations is pick from the 16 stations identified earlier. However, only those stations which have data at the given level are considered.

The interpolated value is calculated by

$$\hat{f}_0' = \sum_{i=1}^{n} a_i \frac{\sigma_0}{\sigma_i} f_i', \qquad n \le 8, \tag{26}$$

where f_i' is the departure from the monthly mean, $\bar{f}_i$, f_i represents the observations, and σ_i is the standard deviation of the observed values from the monthly mean. The coefficients, a_i, are calculated from

$$\sum_{i=1}^{n} a_i \mu_{ij} = \mu_{0j}, \qquad j=1,\ldots,n \tag{27}$$

where μ_{ij} is the correlation coefficient between the j-th and i-th stations. It should be noted that the calculated correlation coefficients μ_{ij} include observational errors.

The residuals used in the horizontal check are defined by

$$\delta f_0 = f_0' - \hat{f}_0' , \tag{28}$$

and the allowable residuals are defined by

$$\Delta f_0 = N_f \cdot \varepsilon \cdot \sigma_0 , \tag{29}$$

where

$$\varepsilon = \sqrt{1-\sum_{i=1}^{n} a_i \cdot \mu_{0j}} . \tag{30}$$

The quality of the results using optimal interpolation for upper-air data depends on the quality of the atmospheric statistics used in the interpolation. Errors in the mean values, standard deviations, and correlation functions lead to errors in the interpolation.

In the current version of the CQC, a multi-year global set of monthly means and standard deviations at mandatory levels from 1000 to 10 hPa on a 5x10 degree latitude-longitude grid was used. Two datasets have been used in producing this composite dataset: a climatology calculated at the RIHMI-WDCB (Russia) by Reitenbach and Sterin (1987), and a climatology calculated at the National Climatic

Data Center by Eric Gadberry in 1993 from ECMWF forecasts for the 1978-1990 period on a 2.5 x 2.5 latitude-longitude grid. Both datasets have strengths and weaknesses. The datasets have been combined to yield a more complete dataset. For each radiosonde station, the mean and standard deviation were obtained by an interpolation from the four nearest gridpoints.

The correlation function for geopotential height and temperature, used in (26), at mandatory levels is calculated from

$$\mu(r) = c e^{-a \cdot r}(\cos br + \frac{a}{b} \sin br) \tag{31}$$

where r is the distance between stations and is in thousands of kilometers. For geopotential height the constants are: a = 0.658, b = 1.033 and c = 0.986. For temperature the constants are, a = 0.807, b = 1.190 and c = 0.893 (Alduchov and Reitenbach, 1990). Currently, the 500 hPa correlation function is used for all levels. This will be corrected in the future. Experiments show that using the correct correlation functions at each mandatory level gives more accurate estimates of the allowable residuals (29) and (30). The same correlation function used for geopotential height is used for geopotential thickness.

Horizontal optimal interpolation of wind components must take into account the vector nature of the wind. Both wind components are used to produce interpolated wind component values. Hence, the correlation functions for each wind component and a cross correlation function between U and V components are needed. Unfortunately, a dataset with accurate wind component correlation functions does not exist. Instead, a set of geostrophic correlation functions are calculated using the geopotential height correlation (31). The geostrophic correlation functions are:

$$\mu_{uu}(r,\alpha) = c e^{-ar} (\cos br \sin^2\alpha - \frac{\sin br}{b} (a - \frac{\cos^2\alpha}{r})) \tag{32}$$

$$\mu_{vv}(r,\alpha) = c e^{-ar} (\cos br \cos^2\alpha - \frac{\sin br}{b} (a - \frac{\sin^2\alpha}{r})) \tag{33}$$

$$\mu_{uv}(r,\alpha) = c e^{-ar}(\cos br - \frac{\sin br}{b}(a+\frac{1}{r}))\sin\alpha\cos\alpha \qquad (34)$$

where α is the angle between the vector connecting the two radiosonde stations and the x-axis which is oriented west to east; and a, b, c are constants determined for the same geopotential height.

The correlation function for dewpoint depression is

$$\mu(r) = 0.9\ e^{-0.98r}. \qquad (35)$$

Figs. 3 through 14 evaluate the horizontal interpolation using a global set of upper-air stations from 0 UTC January 15, 1989. These figures show the accuracy of the interpolation in terms of residuals and allowable residuals.

How close to zero are the mean values of the actual residuals? Two factors affect the mean value of the residuals.

1) The accuracy of the mean monthly values (climatology), and

2) the presence of systematic errors.

Figs. 3,5,7,9,11, and 13 show that the mean values of the residuals are generally close to zero which implies that mean monthly values are accurate and there are no large systematic errors in the data.

The rms values of the actual residuals show how accurately the data can be predicted using horizontal interpolation (e.g., what are the magnitudes of the random errors in the interpolation). For comparison, these figures show the rms value of the residual of the observed and climatic values. Figs. 3, 7, 9, 11, and 13 show that horizontal optimal interpolated rms values are much smaller than the climatic residuals. The difference is 3 to 4 times smaller for geopotential heights (best case), and only 1.2 to 1.3 times smaller for humidity data (the least horizontally correlated variable).

The accuracy of the standard deviations and correlation functions is seen by comparing the rms values of the actual residuals to values of the allowable residuals (see Figs. 3, 5, 7, and 9). A small difference in the rms of the actual residuals and the allowable residuals implies that accurate statistics were used in the optimal interpolation. The figures show that there is very good agreement for geopotential height between actual and allowable residuals, good agreement for temperature and the U component of the wind, and fairly poor agreement for the V component of the wind and the dewpoint depression.

The vertical distribution of the rms values of the standard deviations of the climatic data, which are used to calculate the allowable residuals are shown in Figs. 11 and 13. The figures show that the differences between the rms values of the actual and allowable residuals are very similar in magnitude to the differences between the rms of the climatic residuals and the standard deviations of the climatic data. This means that the standard deviations for both these elements (V-wind and dewpoint depression) are inaccurate, and do not accurately represent the real data. One way to improve the results of the horizontal optimal interpolation is to calculate climatic means and standard deviations with greater accuracy.

Another parameter which depicts the accuracy of the horizontal interpolation is the distribution of actual residuals normalized by the allowable residuals. These distributions are shown in Figs. 1, 4, 6, 8, 10, and 12. The figures show that virtually all the actual residual to allowable residuals are in the range of -4.0 and 4.0. So, we can conclude that $N_f = 4$ is correct.

2.3 Vertical optimal interpolation of upper-air variables

Vertical optimal interpolation is similar to horizontal optimal interpolation. The only difference is the technique for choosing the influencing data points and how the correlation coefficients are defined. For each observed datum at any interior mandatory level, data are interpolated from the nearest upper and lower mandatory levels

$$\hat{f}'_i = a^i_{i+1}\frac{\sigma_i}{\sigma_{i-1}}f'_{i-1} + a^i_{i+1}\frac{\sigma_i}{\sigma_{i+1}}f'_{i+1}, \quad i=2,\ldots,n-1, \tag{36}$$

$$\hat{f}'_1 = a^1_2\frac{\sigma_1}{\sigma_2}f'_2; \quad \hat{f}'_n = a^n_{n-1}\frac{\sigma_n}{\sigma_{n-1}}f'_{n-1}; \tag{37}$$

where f'_i is the deviation of observed value f_i from the climatic monthly mean, $\bar{f}_i$, at the i-th mandatory level, σ_i is the climatic standard deviation, and a^i_j are the coefficients of vertical optimal interpolation.

For interior levels, the coefficients, a^i_j, are found by solving the following system of linear equations

$$\begin{aligned} a^i_{i-1} + \mu^{i+1}_{i-1}a^i_{i+1} &= \mu^i_{i-1} , \\ \mu^{i+1}_{i-1}a^i_{i-1} + a^i_{i-1} &= \mu^i_{i+1} . \end{aligned} \tag{38}$$

The coefficients are defined at the top and bottom mandatory levels by

$$\begin{aligned} a^1_2 &= \mu^1_2, \\ a^n_{n-1} &= \mu^n_{n-1} , \end{aligned} \tag{39}$$

where μ^k_i is the correlation coefficient between observed values at the k-th and i-th mandatory levels.

Actual residuals of the vertical quality control check are defined as

$$\delta f_i = f'_i - \hat{f}'_i, \tag{40}$$

and the allowable residuals are

$$\Delta f_i = N_f \cdot \varepsilon \cdot \sigma_i , \tag{41}$$

where

$$\varepsilon = \sqrt{1. - a^i_{i-1}\,\mu^i_{i-1} - a^i_{i+1}\,\mu^i_{i-1}} \tag{42}$$

for interior levels and

$$\varepsilon = \sqrt{1. - a_2^1\ \mu_2^1}; \quad \varepsilon = \sqrt{1. - a_{n-1}^n\ \mu_{n-1}^n}; \tag{43}$$

at the top and bottom mandatory levels.

Vertical optimal interpolation can be performed using climatic monthly means, standard deviations, and the set of vertical correlation coefficients between mandatory levels. If the coefficients of vertical correlation between mandatory levels are known, vertical optimal interpolation can be performed using these statistics. A set of correlation matrices are calculated using observed values for different seasons and different latitudinal zones. Correlation matrices for all upper-air variables and all zones for winter are shown in Tables 4 through 8.

Figs. 15 through 24 show the accuracy of vertical optimal interpolation for the various upper-air variables. These figures show that as a rule vertical optimal interpolation is more accurate than horizontal optimal interpolation. The greater accuracy of vertical optimal interpolation arises because observed values at different levels of a vertical profile have a greater correlation than the correlation between the data at the same levels between radiosonde stations. Moreover, there is a very important role for vertical interpolation to play in determining observational errors. In the case of horizontal interpolation, observational errors at different stations are independent (or may be weakly correlated for the same model of radiosonde). However, observational errors are very highly correlated in vertical interpolation, especially for geopotential height and wind. This correlation is due to the calculation procedures used at upper-air stations for geopotential heights and winds.

Vertical interpolation is more accurate than horizontal interpolation for interior mandatory levels. The exception is dewpoint depression. Poorer results are obtained at the bottom and top mandatory levels. This is because extrapolation is performed at these levels. Vertical interpolation of dewpoint depression does not yield accurate results; however, it is somewhat better than horizontal interpolation. This result is due to the following: first, humidity is not well correlated vertically in comparison with the other upper-air variables; second, atmospheric statistics are not very accurate for humidity. For variables other than humidity, vertical optimal interpolation accurately interpolates from the data and the accuracy of this interpolation is known (this can be seen in comparing the rms values of the actual and allowable residuals and the distributions of ratios between actual and allowable residuals, shown in Figs. 16, 18, 20, 22, 24).

3. Geostrophic relationship

The geostrophic wind check is based on the assumption that the geostrophic wind should be close to the real wind in the free atmosphere. Interpolated wind values are made at each mandatory level for each station by use of the optimal differentiation method with data from surrounding stations (Gandin and Kagan, 1976). The zonal and meridional components of the geostrophic wind at a station are calculated by

$$\begin{aligned} \hat{u}'_{gi} &= \sum_{j=1}^{n} a_j \frac{\sigma_{ugi}}{\sigma_{Hij}} H'_{ij} , \\ \hat{v}'_{gi} &= \sum_{j=1}^{n} b_j \frac{\sigma_{vgi}}{\sigma_{Hij}} H'_{ij} , \\ & \qquad j = 1, \ldots, n ; \end{aligned} \tag{44}$$

where H'_{ij} is the departure of geopotential height from the monthly mean at i-th level and j-th stations, σ_{Hij} is the standard deviation of geopotential height, σ_{ugi} and σ_{vgi} are the standard deviations of the zonal and meridional components of geostrophic wind at i-th level, and a_j and b_j are the calculated coefficients for each station.

Actual residuals of the geostrophic wind are defined by

$$\begin{aligned} \delta u_{gi} &= u'_i - \hat{u}'_{gi} \\ \delta v_{gi} &= v'_i - \hat{v}'_{gi} \end{aligned} \tag{45}$$

where u'_i and v'_i are the departures of the real wind from the monthly means.

a_j and b_j are calculated as solutions of the system of linear equations

$$\sum_{j=1}^{n} a_j\mu_{kj} + a_k = \mu_{ugk} ,$$
$$\sum_{j=1}^{n} b_j\mu_{kj} + b_k = \mu_{vgk} , \qquad (46)$$
$$for\ k = 1,\ldots,n .$$

In (46) μ_{kj} is the correlation coefficient between observed geopotential height data at i-th level of the k-th and j-th stations, μ_{ugk} and μ_{vgk} are the correlation coefficients between the geopotential height at the k-th station and the zonal and meridional components of geostrophic wind at the station being controlled.

The allowable residuals of the geostrophic check are defined as

$$\Delta u_{gi} = \sqrt{1-\sum_{j=1}^{n} a_j\mu_{ugi}} \cdot \sigma_{ugi} ,$$
$$\Delta v_{gi} = \sqrt{1-\sum_{j=1}^{n} b_j\mu_{vgi}} \cdot \sigma_{vgi} . \qquad (47)$$

Two very important problems must be addressed in order to use the geostrophic interpolation. First, the correlation structure of the geostrophic wind must be defined; and second, the method of selecting neighboring stations for each case must be determine. In this paper, the correlation structure is derived from the correlation structure of geopotential height using the geostrophic equations (49). To describe the geopotential height spatial correlation structure, the correlation function

$$\mu(r) = ce^{-ar}(\cos br + \frac{a}{b}\sin br) \qquad (48)$$

is used. In (48) a=0.658,b=1.033,c=0.986 (Alduchov and Reitenbach, 1990).

The geostrophic equations are

$$u_g = -A\frac{\partial H}{\partial y}, \qquad v_g = A\frac{\partial H}{\partial x}, \tag{49}$$

where

$$A = \frac{g}{2\omega\sin\varphi}, \tag{50}$$

g is the acceleration of gravity (=9.81 m/s), ω is the angular velocity of the Earth's rotation (=7.29 10^{-5} sec^{-1}), and φ is latitude. The correlation function is assumed to be homogeneous, then from (48) and (49)

$$\mu_{ug}(r,\alpha) = -ce^{-ar}\frac{\sin br}{b}\sin\alpha, \qquad \mu_{vg}(r,\alpha) = ce^{-ar}\frac{\sin br}{b}\cos\alpha. \tag{51}$$

The correlation between geopotential height and the zonal and meridional components of the wind is given by (51), where α is the angle of the vector defined by the two stations and the x-axis (west-east), and **r** is the distance between the two points.

Because the coefficient, A, in (49) is not defined when the latitude φ is close to zero, the geostrophic equations are used only below -20° and above +20° latitude.

Standard deviations (Gandin and Kagan, 1976) are calculated from

$$m_{ug}(r,\alpha) = cA^2(a^2+b^2)\left(\cos br\sin^2\alpha - \frac{\sin br}{b}\left(a-\frac{\cos^2\alpha}{r}\right)\right)\sigma_H^2, \qquad m_{vg}(r,\alpha) = cA^2(a^2+b^2)\left(\cos br\cos^2\alpha - \frac{\sin br}{b}\left(a-\frac{\sin^2\alpha}{r}\right)\right)\sigma_H^2, \tag{52}$$

Taking the limit of (52) for $r \rightarrow 0$) yields the following relationship

$$\sigma_{ugi} = \sigma_{vgi} = |A|\sqrt{a^2+b^2}\,\sigma_{Hi} \tag{53}$$

where σ_{Hi} is the standard deviation of geopotential height.

It should be noted that equations (48) through (53) are theoretical estimates made with numerous assumptions. These equations are used after making adjustments based on experience.

The standard deviations for the zonal and meridional components of the geostrophic wind are modified to

$$\sigma_{ugi} = \sigma_{vgi} = |A| \; 0.95 \; \sigma_{Hi}. \tag{54}$$

Better results are obtained using this modified equation than (53).

To select eight neighboring, influencing stations to provide a geostrophic wind calculations, a procedure is used which is similar to the one for horizontal interpolation of geopotential height and temperature. Influencing stations must be located within 2000 km of the given station. To find an even and symmetrically station distribution, a circle of influence (2000 km radius) is constructed with eight 45 degrees sectors. In each sector, no more than two stations are picked. The stations picked have the greatest correlation of geopotential height with geostrophic wind at the given station. Any of these 16, or less, surrounding stations can be used in the geostrophic wind calculations. The method of choosing the eight stations, which will be use in (44) is:

a) in each of the eight sectors, the station which has the largest correlation of geopotential height to geostrophic wind data is found using (51);

b) if a sector is empty, then a station is chosen from a sector which contains more than one station;

c) if the number of designated "influencing" stations is less than or equal to eight, then all candidate stations are using in the calculations;

d) with the chosen set of stations, weights a_j and b_j are calculated from (46) for use in (44). Actual and allowable residuals are calculated at those levels using observations from each influencing station;

e) if the station data being tested has data which is missing at one or more of the influencing stations, then for this level a new set of eight influencing stations is picked from the 16 stations identified earlier.

Figs. 25 through 28 show the results of using this technique to calculate the geostrophic wind. These figures show that, for the U component below 100 hPa and the V component below 300 hPa, there is good agreement between actual and allowable residuals. Above these levels the agreement is not very good. This means that the calculated statistical structure of the wind does not quite correspondent to the real structure of the geostrophic wind. More accurate correlation functions of geopotential and the correlation function for geostrophic wind at different heights should improve the calculations of geostrophic winds and our estimates of accuracy for these calculations. But, using the geostrophic approximations of the real wind doesn't result in a significant reduction of the residuals in comparison with a climatic approximation (climatic check). The conclusion is that the geostrophic wind check can be helpful in some cases (because any additional information is helpful in decision making), but this check <u>cannot</u> be the basis for making decisions about errors in wind observations.

4. Thermal relationships

The thermal wind check is based on the assumption that the geostrophic wind is close to the real wind in the free atmosphere and variations of the geostrophic wind and the real wind with height are defined by the thermal wind. The thermal wind is calculated at each station for layers between mandatory levels by optimal differentiation (Gandin and Kagan,1976) using temperature data from neighboring upper-air stations.

Differentiating the geostrophic wind equations (49) with respect to ln(p) and use hydrostatic equation yields

$$\begin{aligned} \frac{\partial u_g}{\partial \ln p} &= -B\frac{\partial T}{\partial y}\ , \\ \frac{\partial v_g}{\partial \ln p} &= B\frac{\partial T}{\partial x}\ , \end{aligned} \tag{55}$$

where

$$B = \frac{R}{2\omega \sin\varphi}, \tag{56}$$

R is the gas constant for dry air (=287 m^2/sec^2 °K), ω is the angular velocity of Earth's rotation, and φ is the latitude.

Integrating (55) yields the following relationship for the variation of geostrophic wind with height

$$\begin{aligned} u_{ti}^{i+1} &= \Delta u_{gi}^{i+1} = B\frac{\partial T_i^{i+1}}{\partial y}\ln\frac{p_i}{pi+1}, \\ v_{ti}^{i+1} &= \Delta v_{gi}^{i+1} = -B\frac{\partial T_i^{i+1}}{\partial x}\ln\frac{p_i}{p_{i+1}}, \end{aligned} \tag{57}$$

where u_{ti} and v_{ti} are the variations of the zonal and meridional components of geostrophic wind between mandatory pressure levels p_i and p_{i+1}; T_i^{i+1} is the averaged temperature of the layer.

From (57), the variations of zonal and meridional wind with heights are calculated

$$\begin{aligned} \hat{u}_{ti}^{i+1'} &= (\hat{u}'_{ti+1}/\sigma_{uti+1} + \hat{u}'_{ti}/\sigma_{uti})\,\sigma_{u_{ti}^{i+1}}/2, \\ \hat{v}_{ti}^{i+1'} &= (\hat{v}'_{ti+1}/\sigma_{vti+1} + \hat{v}'_{ti}/\sigma_{uti})\,\sigma_{v_{ti}^{i+1}}/2, \end{aligned} \tag{58}$$

where $\hat{u}_{ti}$ and $\hat{v}_{ti}$ are linear combinations defined by

$$\begin{aligned} \hat{u}'_{ti} &= \sum_{j=1}^{n} a_j \frac{\sigma_{uti}}{\sigma_{Tij}} T'_{ij}, \\ \hat{v}'_{ti} &= \sum_{j=1}^{n} b_j \frac{\sigma_{vti}}{\sigma_{Tij}} T'_{ij}, \end{aligned} \tag{59}$$

and T'_{ij} is the departure of the temperature from the monthly mean at the i-th level and j-th station, σ_{Tij} is the standard deviation of geopotential, σ_{uti} and σ_{vti} are the standard deviations of the zonal and meridional components of the geostrophic wind at the i-th level, a_j and b_j are the calculated coefficients from the neighboring stations,

and

$$\sigma_{u_i^{i+1}}=\sqrt{\sigma_{uti}^2+\sigma_{uti+1}^2+2\mu_{T_{ii+1}}\ \sigma_{uti}\sigma_{u_{ti+1}}}/2\ ,$$
$$\sigma_{v_i^{i+1}}=\sqrt{\sigma_{vti}^2+\sigma_{vti+1}^2+2\mu_T^{ii+1}\ \sigma_{vti}\sigma_{v_{ti+1}}}/2\ , \tag{60}$$

where μ_{ij} - correlation coefficient between temperatures at i-th and j-th levels. Actual geostrophic residuals are defined by the differences

$$\delta u_{ti} = u'_{i+1} - u'_i - \hat{u}_{ti}^{i+1\prime}\ ,$$
$$\delta v_{gi} = v'_{i+1} - v'_i - \hat{v}_{ti}^{i+1\prime}\ , \tag{61}$$

where u'_i and v'_i are the departures of real wind components from monthly means.

Coefficients a_j and b_j in (59) are calculated as a solution of the system of linear equations

$$\sum_{j=1}^{n} a_j\mu_{kj} + a_k = \mu_{ugk}$$
$$\sum_{j=1}^{n} b_j\mu_{kj} + b_k = \mu_{vgk} \tag{62}$$

where μ_{kj} is the correlation coefficient between observed values of temperature at i-th level of k-th and j-th stations, μ_{utk} and μ_{vtk} are the correlation coefficients between temperatures at k-th station and the zonal and meridional components of the thermal wind.

Allowable thermal wind residuals are defined as

$$\Delta u_{ti} = \sqrt{1-\sum_{j=1}^{n} a_j\mu_{uti}}\ \sigma_{u_{ti}^{i+1}}\ ,$$
$$\Delta v_{ti} = \sqrt{1-\sum_{j=1}^{n} b_j\mu_{vti}}\ \sigma_{v_{ti}^{i+1}}\ . \tag{63}$$

The correlation structure of the thermal wind is defined from the correlation structure of temperature using derived geostrophic relationships.

To describe the spacial temperature correlation structure function, the correlation function

$$\mu(r) = ce^{-ar}(\cos br + \frac{a}{b}\sin br) \tag{64}$$

is used where a=0.658, b=1.033, c=0.986 (Alduchov and Reitenbach,1991).

If the correlation function is assumed to be homogeneous, then (57) and (64) yield the following correlation functions

$$\begin{aligned} \mu_{tu}(r,\alpha) &= ce^{-ar}\frac{\sin br}{b}\sin\alpha , \\ \mu_{tv}(r,\alpha) &= -ce^{-ar}\frac{\sin br}{b}\cos\alpha , \end{aligned} \tag{65}$$

for the correlation between temperature and the zonal and meridional components of the wind. In (65), α is the angle between the line connecting the two stations (points) and the x-axis and r is the distance between the two stations.

The coefficient B, (56), doesn't make sense when latitude ϕ is close to zero, hence the geostrophic check is used only below -20° and above +20° latitude.

The standard deviations of the thermal wind components (using the covariance functions from the temperature covariance and the thermal wind relationship (similar to the procedure for geostrophic wind) and taking the limit r → 0) are:

$$\sigma_{uti} = \sigma_{vti} = |B| \ln\frac{p_i}{p_{i+1}}\sqrt{a^2+b^2}\ \sigma_{Ti} , \tag{66}$$

where σ_{Ti} - standard deviations for geopotential.

The procedure by which stations are selected for use in the interpolation is very similar to the ones for horizontal and geostrophic interpolation. Influencing stations must be within 2000 km of the station whose data are being checked. To pick an even and symmetrically distribution of stations, a circle of influence (2000 km radius) with eight 45 degrees sectors is constructed. In each sector, two stations are selected which have the largest correlation between thermal wind values at test station and temperature values with the interpolating station. Any of the 16 (or less) surrounding stations can be used as influencing stations in calculations of the thermal wind.

The method of choosing stations for use in (62) is the same as used in the geostrophic check.

Figs. 29 and 30 show the magnitude of the errors that can be detected using the thermal wind relationships. It can be seen that for U and V wind components below the 150 hPa level there is good agreement between the actual and allowable residuals, but above these levels, the agreement is not very good. This result is almost identical to the results using checks based on the geostrophic wind. It confirms that the statistical structure of the wind contains errors. The statistical structure being used does not quite correspond to the real structure of the geostrophic wind and thermal wind, at least at high levels. In our opinion, using different correlation functions for the geopotential and temperature (and hence, different correlation functions for geostrophic and thermal wind) for different heights should improve the accuracy of geostrophic and thermal wind calculations and our estimates of the accuracy of these calculations. In the calculation of the thermal wind, vertical correlations of temperature are used and these correlations need to be determined more accurately. Figs. 29 and 30 show that the thermal wind approximation to the real wind yields, on average, standard deviations of differences between interpolated and real variations of wind between adjacent mandatory levels, of approximately 5 m/s. In the middle troposphere, it reduces by 50% the standard deviations in comparison with the climatic check. At other levels, the improvement is not as significant compared to the climatic check. Figs. 31 and 32 show the distribution of the normalized actual residuals for the thermal wind approximation.

The thermal check is useful in the troposphere, but this check cannot be the primary one used in decision making about errors in wind observations.

5. Linear interpolation from significant levels to mandatory levels

To interpolate data from significant levels to mandatory levels we use linear interpolation. Linear interpolation can be used because of the definition of significant levels and criteria for choosing significant levels in a sounding. As a rule, significant levels are identified as those levels which enable one to reproduce a sounding of a upper-air variable by linear interpolation with an accuracy up to 1.0-2.0 °C for temperature, 10-15 % - for relative humidity and 10-15 degrees for wind direction, and about 5 m/s for wind speed (see Federal Meteorological Handbook No.3 (1981) and Instructions to hydrometeorological stations and posts (1976)). These limits for the various upper-air variables enable us to use significant levels to check data at mandatory levels.

The predicted value at a mandatory level, pressure p_i, is calculated from significant level data, pressure p_{s1} and p_{s2}, from

$$\hat{f}_i = \frac{a_2 \cdot f_{s1} + a_1 \cdot f_{s2}}{a_1 + a_2} , \qquad (67)$$

where the coefficients a_1 and a_2 are proportional to the distance between significant levels p_{s1} and p_{s2} and mandatory level p_i, respectively. These coefficients are calculated by

$$a_j = \frac{R}{g} \cdot T_s \cdot \left| \ln \frac{p_{sj}}{p_i} \right| , \qquad (68)$$

where R is the specific gas constant for dry air, g is the gravitational constant, and T_s is the averaged temperature (in °K) for the layer between P_i and P_{i+1}.

Significant levels p_{s1} and p_{s2} are selected below and above a mandatory level p_i such that the distance between the two significant levels (a_1+a_2) is less than 6 km.

Figs. 33, 34, 35, and 36 show the results of interpolating significant levels to mandatory for temperature, the U and V wind components, and dewpoint depression at mandatory levels. Figs. 37, 38, 39, and 40 show the distribution of departure of the observed values from the interpolated values. The accuracy of the interpolation from significant to mandatory levels has a weak dependency on height and is the <u>best</u>

method for checking mandatory level data. The weak dependency of the residuals on height enable us to use constant allowable residuals for every upper-air variable. In the current version of CQC, the following allowable residuals are used: 3 °C for temperature, 5 m/s for the U and V wind components, and 5 °C for dewpoint depression. These values were used for the allowable residuals to preparing figs. 37, 38, 39, and 40.

It would be very nice to have the quality control check of mandatory levels based on the interpolation of data from significant levels. But, unfortunately only about 30 to 60% of the temperature, wind, and humidity data at mandatory levels have adjacent significant level data. Data at significant levels do not usually contain geopotential height datum, and data at significant levels contain errors just like the mandatory level data. Therefore it is impossible to make this method of quality control check the primary method in the CQC procedures. But, it is too powerful a method not to be used in the CQC of upper-air data.

6. Linear interpolation from mandatory levels to significant levels

Data at significant levels may contain errors and must be checked. This is accomplished using the already checked data from mandatory levels.

Experience with data at mandatory levels shows that it is possible to detect errors with magnitude from 5 to 10 °C in temperature, 30 to 60 gpm in geopotential height, 10 to 15 °C in dewpoint depression, and 10 to 20 m/s in the zonal and meridional wind components. This is accomplished using accurate interpolation methods and information redundancy in the sounding data to quality control the data at points on a grid formed by observations on isobaric surfaces. The scale of the grid is 300 to 500 km in the horizontal (a typical distance between neighboring upper-air stations) and 1.5 to 3.0 km in the vertical (a typical interval between mandatory levels surfaces). With the data at mandatory levels being checked with this accuracy, it is impossible to determine the vertical profile of the corresponding upper-air variable with a higher accuracy than these error ranges. The accuracy of the whole vertical profile is determined by the lowest accuracy of all of the upper-air variables used to construct the profile. Thus, optimum accuracy in the quality control check of single data points is accomplished by controlling the accuracy of data at mandatory levels.

To reach the above mentioned accuracy in checking upper-air variables at mandatory levels, rather complicated interpolation and decision-making methods must be used. To check the same variables at significant levels with about the same accuracy using data already checked at mandatory levels, it is sufficient to use interpolation (70) and simple decision-making algorithms. This is due to the fact that the interpolation distance using mandatory levels to significant levels is less than half the interpolation distance between mandatory levels.

The quality control method for data at significant levels is as follows: the value f_o is compared with the result of a linear interpolation of values from the two closest mandatory levels given by:

$$\hat{f}_o = a_i \cdot f_i + a_{i+1} \cdot f_{i+1} , \tag{69}$$

where f_i and f_{i+1} are CQCed values from the i-th and i+1-th mandatory levels, a_i and a_{i+1} are linear interpolation coefficients defined by

$$a_i = \ln\frac{p_o}{p_{i+1}} / \ln\frac{p_i}{p_{i+1}} , \qquad a_{i+1} = \ln\frac{p_i}{p_o} / \ln\frac{p_i}{p_{i+1}} , \tag{70}$$

and p_o is the pressure at the significant level, p_i and p_{i+1} are the pressures at the mandatory levels.

The actual residual of this control method at the significant level is defined by

$$\delta f_o = \hat{f}_o - f_o , \tag{71}$$

It could be assumed that if the absolute value of actual residual (71) is large (i.e. difference between the observed and interpolated value), then the observed is erroneous and must be rejected. On the other hand, if absolute value of (71) is small the observed value is correct.

The problem is developing a criterion to determine whether the actual residual is large or small.

To define this criterion consider the actual residual(71) and represent it as follows

$$\begin{aligned} \delta f_o &= a_i \cdot f_i + a_{i+1} \cdot f_{i+1} - f_o = \\ &(a_i \cdot f_i' + a_{i+1} \cdot f_{i+1}' - f') + \\ &(a_i \cdot \overline{f}_i + a_{i+1} \cdot \overline{f}_{i+1} - \overline{f}) \ , \end{aligned} \tag{72}$$

where $\overline{f}$ is the mean monthly value of f at the corresponding level and f′ is the deviation of the observed value from the monthly mean. Thus (72) becomes

$$f = \overline{f} + f' \ , \tag{73}$$

The second part of (72) is close to zero and it can be ignored in further computations. It is assumed that within the layer between adjacent mandatory levels mean values of upper-air variables vary in a linear manner.

Let's consider now the mean square of the actual residual

$$\begin{aligned} \overline{\delta f_o^2} &= \overline{f_o'^2} - 2 \cdot a_i \cdot \overline{f_o' \cdot f_i'} - 2 \cdot a_{i+1} \cdot \overline{f_o' \cdot f_{i+1}'} + \\ &a_i^2 \cdot \overline{f_i'^2} + a_{i+1}^2 \cdot \overline{f_{i+1}'}^2 + 2 \cdot a_i \cdot a_{i+1} \cdot \overline{f_i' \cdot f_{i+1}'} \ , \end{aligned} \tag{74}$$

and use

$$\overline{f_a' \cdot f_b'} = \frac{\overline{f_a' \cdot f_b'}}{\sigma_a \cdot \sigma_b} \cdot \sigma_a \cdot \sigma_b = \mu_{ab} \cdot \sigma_a \cdot \sigma_b \ , \tag{75}$$

where σ_a and σ_b are the standard deviations, and μ_{ab} is the correlation coefficient of parameter f at points a and b.

Hence (74) can be written as

$$\begin{aligned} \overline{\delta f_o^2} &= \sigma_o^2 - 2 \cdot a_i \cdot \mu_{oi} \cdot \sigma_o \cdot \sigma_i - 2 \cdot a_{i+1} \cdot \mu_{oi+1} \cdot \sigma_o \cdot \sigma_{i+1} + \\ &a_i^2 \cdot \sigma_i^2 + a_{i+1}^2 \cdot \sigma_{i+1}^2 + 2 \cdot a_i \cdot a_{i+1} \cdot \mu_{ii+1} \cdot \sigma_i \cdot \sigma_{i+1} \ , \end{aligned} \tag{76}$$

Assuming that the standard deviations at adjacent mandatory levels and at any significant level between them are approximately equal means

$$\sigma_i = \sigma_{i+1} = \sigma_o = \sigma_c \tag{77}$$

Also assuming that

$$\mu_{oi} = \mu_{ii+1} + a_i \cdot (1 - \mu_{ii+1}) \ , \tag{78}$$

Equation (78) means that the correlation coefficient between observed values at significant levels and adjacent mandatory levels changes according to a linear law within the limits from 1 to μ_{ii+1}, where μ_{ii+1} is the correlation coefficient between the i-th and i+1-th mandatory levels.

Then (76) becomes

$$E^2 = \overline{\delta f_o^2} = 2 \cdot a_i \cdot a_{i+1} \cdot (1 - \mu_{ii+1}) \cdot \sigma_c^2 \ . \tag{79}$$

Thus, in each case it becomes possible to estimate the allowable residuals by

$$\Delta f_o = N_f \cdot E = N_f \cdot \sqrt{2 \cdot a_i \cdot a_{i+1} \cdot (1-\mu_{ii+1})} \cdot \sigma_c \ , \tag{80}$$

When the value of Δ exceeds $|\delta f_o|$ this indicates there is likely an error in the observed value f_o.

In summary, it should be noted the following four assumptions have been made in checking the data. First, it is assumed that the mean values of upper-air variables at significant levels can be estimated from the mean values of the same variable at adjacent mandatory levels using linear relationships

$$\overline{f_o} = a_i \cdot \overline{f_i} + a_{i+1} \cdot \overline{f_{i+1}} \ , \tag{81}$$

In (81) a_i and a_{i+1} coefficients are defined in (70). It is clear that the closer the mandatory levels to the significant level, the better the assumption. Second, it is assumed that standard deviations at the significant level and adjacent mandatory levels (77) are equal. This assumption is quite reasonable especially when σ_c is the averaged value

$$\sigma_c = \sigma_o = (\sigma_i + \sigma_{i+1})/2 \; , \tag{82}$$

or even

$$\sigma_c = \sigma_o = a_i \cdot \sigma_i + a_{i+1} \cdot \sigma_{i+1} \; . \tag{83}$$

Third, it is assumed that correlation coefficients between the observed value at the given significant level and those at the adjacent mandatory levels are defined by (78). For small values of the correlation function variable, the correlation function is proportional to the argument squared.

$$\mu(\rho) = 1 - c_1 \cdot \rho^2 \tag{84}$$

where ρ represents distance. When ρ is not small, the correlation coefficient is proportional to ρ (a well known "first degree law")

$$\mu(\rho) = 1 - c_2 \cdot \rho \; , \tag{85}$$

the use of equation (78) for mandatory levels close to the significant level will lead to relatively small errors in estimating the correlation coefficient. If the surfaces are close so that ρ is very small, the use of equation (78) will lead to an underestimation of the correlation coefficients. This is useful, as it allows us to make use of a characteristic of significant levels and the underestimated correlation coefficients to extend the limits (gate) for the value being checked at this point. The use of a wide gate for this data point maybe more correct than a narrower gate.

Finally, it is assumed that the actual residuals are normally distributed with a mean value of zero. This assumption is generally justified though the distribution density of the actual residuals of different variables parameters being controlled can differ from the normal distribution. However, these differences can be considered with the help of coefficient variation N_f in (80).

Thus, the successful use of this method is dependent on the accuracy of assumptions. However, with regard to upper-air data, experience shows that this control method yields rather good results. Figs. 41 through 48 show the distribution of the actual and admissible differences of the present control method. These figures show that the theoretical estimation (allowable residuals) of the differences between the interpolated value (from mandatory levels to a significant level) and the observation at this level is, on the average, in good agreement with the actual residuals.

III. Decision making algorithms

A decision making algorithm (DMA) decides whether an observed value is correct or erroneous. The DMA is constructed based on an analysis of each CQC component response to possible errors in each observed variable.

1. DMA for geopotential height and temperature at mandatory levels

The following types of errors are possible in geopotential and temperature observed values at the mandatory levels:

i. garbled geopotential height at a lower, intermediate, and upper level;

ii. garbled temperature at a lower, intermediate, and upper level;

iii. miscalculation of the thickness between adjacent mandatory levels resulting in erroneous geopotential height above this layer by a constant value;

iv. radiosonde malfunction starting in a lower or intermediate level producing erroneous temperatures and corresponding erroneous geopotential heights which do not violate the hydrostatic equation;

v. an error in the station identifier (usually WMO number) or station coordinates (i.e. the sounding is assigned to wrong point of the globe);

vi. wrong coding and/or complete garbling of the sounding, as well as combinations of the above-mentioned errors.

For each type of error, the actual residuals which will be produced by a specified error can be estimated. Table 9 shows these estimates. This table is indispensable for creating a decision making algorithm for geopotential and temperature at mandatory levels, since it enables us to solve the inverse task: determine the error given the values of the residuals. However, it is important to recognize that for correct data the actual residuals are not zero. Noise must be included in the actual residual to correctly construct a DMA. Part II of this work shows the "noise" for each component of the CQC.

This makes creating a DMA much more difficult, because it is necessary to distinguish between "noise" and a response to real errors in the data.

To detect and correct errors of type 1, 2, or 3, which often occur in upper-air data, a hydrostatic check is essential. The hydrostatic relationship is the most accurate relations between geopotential height and temperature. Other components of the complex quality control are of secondary importance, used only in those cases, when the hydrostatic check does not lead to a definitive decision. The hydrostatic check does not react to the errors of type 4 and 5. Therefore, the horizontal and vertical checks are the primary checks for these errors. All of the QC checks react to errors of type 6.

The DMA for geopotential and temperature at mandatory levels consists of three logical sections:

a decision making section for suspect values;

error identification and correction section;

section for processing remaining suspect values.

In the first section of the DMA, the following check is carried out if at least one of the actual residual exceeds the corresponding allowable residual:

$$
\begin{aligned}
&|\delta_i^{i+1}| > \Delta_i^{i+1} , \\
&|\delta h_i^H| > \Delta h_i^H , \quad i=1,\ldots,n-1, \\
&|\delta H_i^V| > \Delta H_i^V , \\
&|\delta H_i^H| > \Delta H_i^H , \\
&|\delta T_i^V| > \Delta T_i^V , \\
&|\delta T_i^H| > \Delta T_{i_H} , \\
&|\delta T_i^S| > \Delta T_i^S , \quad i=1,\ldots,n.
\end{aligned}
\tag{86}
$$

where

δ_i^{i+1} and Δ_i^{i+1} - actual (calculated for the specific data point) and allowable residuals of the hydrostatic check,

δh_i^H and Δ_i^H - residuals of the horizontal check of thickness,

δH_i^H and ΔH_i^H - residuals of the horizontal check of geopotential heights,

δT_i^H and ΔT_i^H - residuals for the horizontal check of temperatures,

δH_i^V and ΔH_i^V - residuals for the vertical check of geopotential heights,

δT_i^V and ΔT_i^V - residuals for the vertical check of temperatures,

δT_i^S and ΔT_i^S - residuals for the temperature check using significant level data.

If at least one of these inequalities (86) is true, it is assumed that an error is possible in the sounding and the DMA enters the second section, otherwise the sounding is assumed to be correct and exits out of the DMA.

In the second section of the DMA, the logical analysis of the residuals is carried out for all CQC components from the lowest to the highest level in each sounding.

If an error in T_i or H_i or thickness (errors of the type 1, 2, or 3) is detected, then T_i or H_i and the corresponding actual residuals for all CQC components are recomputed taking into account the corrections made. The sounding is then sent to the first section of the DMA.

Some H_i and T_i values are corrected only after an analysis of the magnitudes and signs of the proposed corrections. Most errors are usually consequence of a garbling of one digit or the sign in the value. In this case, each correction is modified by some admissible value (change in value is limited to one digit or the sign). If such a modification is impossible, it is assumed that more than one symbol is distorted in the calculated value, the proposed correction is made, rounded off to the nearest meter for geopotential height and degree for temperature.

If errors of type 4, 5, or 6 are detected, all of the sounding is considered to be erroneous. No corrections are made and the sounding exits the DMA.

The error identification procedures uses a set of logical variables S, TV, HV, hH, TH, HH and TS which are defined as

$$
\begin{aligned}
S(k,i) &= (|\delta_i^{i+1}| > k\cdot\Delta_i^{i+1}/2)\ , \\
hH(k,i) &= (|\delta h_i^H| > k\cdot\Delta h_i^H/2)\ , \quad i=1,\ldots,n-1, \\
TV(k,i) &= (|\delta T_i^V| > k\cdot\Delta T_i^T/2)\ , \\
TH(k,i) &= (|\delta T_i^H| > k\cdot\Delta T_i^H/2)\ , \\
TS(k,i) &= (|\delta T_i^S| > k\cdot\Delta T_i^S/2)\ , \\
HV(k,i) &= (|\delta H_i^V| > k\cdot\Delta H_i^V/2)\ , \\
HH(k,i) &= (|\delta H_i^H| > k\cdot\Delta H_i^H/2)\ , \quad i=1,\ldots,n;\ k = 1,2.
\end{aligned}
\tag{87}
$$

When the condition is true the value is set to one and zero when false. In (87), the single letter S represents the hydrostatic check. In the first position the character H represents geopotential height, h thickness, and T temperature. In the second position, the character H represents horizontal, V represents vertical, and S represents significant level.

If conditions in (87) are true for k = 1, the error classification is termed "weak" and strong for k = 2.

With this notation, a decision concerning the existence of an error is made according to the following rules and in the following order:

a. Incorrect coding or complete garbling of the sounding.

It is assumed that there is a coding error in the sounding or garbling so that recovery is not possible when

$$\left(\sum_{i=1}^{n-1} S(2,i) \geq \frac{2n}{3}\right) \wedge$$
$$\left(\sum_{i=1}^{n} TH(2,i) \geq \frac{2n+2}{3}\right) \wedge \qquad (88)$$
$$\left(\sum_{i=1}^{n} HH(2,i) \geq \frac{2n+2}{3}\right)$$

is true. In (88), $\wedge$ is the logical "and" symbol. Equation (88) is satisfied when 66.6% of the actual residuals from the hydrostatic check, horizontal check of temperature, and horizontal check of geopotential height exceed the allowable residuals. When this is true the sounding is considered erroneous and analysis of the sounding is halted.

b. Error in the station index or station coordinates.

Identification of errors in station coordinates is based on the fact that reported values differ markedly from climatological values throughout the atmosphere.

The actual residuals from the horizontal and vertical check are defined by (see section 2.1)

$$\delta f = f' - \hat{f}' = f' - \sum_{j=1}^{k} a_j f'_j \ . \qquad (89)$$

where f' represents the departure from the monthly mean and $\hat{f}'$ is the interpolated value of the departure of the monthly mean at the station.

Departures f', from a sounding erroneously assigned to the wrong location have large magnitude and are not balanced by the horizontal interpolated values $\hat{f}'$generated from influencing stations assigned to the correct locations. The interpolated value is defined as $f'_i = a_{i-1}*f'_{i-1} + a_{i+1}*f'_{i+1}$ for vertical check at intermediate levels. The sum of both a-coefficients is close to 1.0. If the values at the i-1-th and i+1-th levels have magnitude D, the interpolated value at the i-th level will have the same value D. Then the observed and interpolated values have magnitude D and the difference between them is zero. A difference of zero means that the actual residuals from a vertical check at intermediate levels will be close to zero.

To obtain the interpolated departure values for the vertical check at the bottom and top levels the relation, $f'_i = a*f'_j$, is used. In this case the coefficient "a" is about 0.7. Hence, the interpolated value has magnitude of about 0.7*D. Hence, difference between observed departure D and interpolated value 0.7*D is about 0.3*D.

It should be noted, that climatological norm in the atmosphere for different regions is usually distinctly different in the lower atmosphere. If this is the case, large values of the actual residuals are expected here.

The hydrostatic check should not react to an error in station location, as the hydrostatic equation is equally true for all regions.

According to these reasons, a decision on an error in station coordinates is made if the following relationship is true

$$\begin{aligned} &\Big(\sum_{i=1}^{n-1} S(2,i) = 0\Big) \wedge \\ &\Big(\sum_{i=1}^{n} HH(1,i) \geq \frac{2n+2}{3}\Big) \wedge \\ &\Big(\sum_{i=1}^{n} TH(i+1) \geq \frac{2n+2}{3}\Big) \wedge \\ &TH(1,1) \wedge TH(1,2) \wedge \\ &HH(1,1) \wedge (\delta T_1^H \cdot \delta T_2^H > 0) \end{aligned} \tag{90}$$

where $\wedge$ is the logical "and". The number of such errors in a sounding should be extremely small.

c. Error τ in T_1 value or error χ in H_1 value.

Error analysis at the first mandatory level is carried out if at least one of the following conditions is true

$$\begin{aligned}&S(2,1) \lor TS(2,1) \lor \\ &((|\delta H_1^H| \geq |\delta H_2^H|) \land hH(1,1)) \lor \\ &((|\delta H_1^V| \geq |\delta H_2^V|) \land HV(1,1)) \lor \\ &((|\delta T_1^H| \geq |\delta T_2^H|) \land TH(1,1)) \lor \\ &((|\delta T_1^V| \geq |\delta T_2^V|) \land TV(1,1))\end{aligned} \tag{91}$$

see (87) for definitions. $\lor$ is the logical "or" symbol.

The most common and simple error is an error in H_1 or T_1 detected by the residuals from the hydrostatic check and the horizontal thickness check or the horizontal geopotential height or the temperature check or vertical geopotential height or temperature check:

$$S(2,1) \land hH(1,1) \land (|\delta H_1^H| \geq |\delta H_2^H|) \land (|\delta h_1^H + \delta_1^2| \leq \Delta_1^2) \tag{92}$$

or

$$S(2,1) \land HH(1,1) \land (|\delta H_1^H| \geq |\delta H_2^H|) \land (|\delta H_1^H + \delta_1^2| \leq \Delta_1^2) \tag{93}$$

or

$$S(2,1) \land TH(1,1) \land (|\delta T_1^H| \geq |\delta T_2^H|) \land (|\delta T_1^H + \delta_1^2/B_1^2|) \leq \Delta_1^2/B_1^2) \tag{94}$$

or

$$S(2,1) \land HV(1,1) \land (|\delta H_1^V| \geq |\delta H_2^V|) \land (|\delta H_1^V + \delta_1^2| \leq \Delta_1^2) \tag{95}$$

or

$$S(2,1) AND TV(1,1) \land (|\delta T_1^V| \geq |\delta T_2^V|) \land (|\delta T_1^V - \delta_1^2/B_1^2| \leq \Delta_1^2/B_1^2) \tag{96}$$

If one of the conditions (92) to (96) is true, the error in H_i or T_i is determined by the hydrostatic residual check

$$\chi_1 = -\delta_1^2 \tag{97}$$

or

$$\tau_1 = -\delta_1^2/B_1^2 \ , \tag{98}$$

respectively.

It is possible that T_1 and H_1 and/or H_2 are garbled. The residual of the hydrostatic .eck, δ_1^2, is due to two or three errors and cannot be used in determining the error in T_1. Only the vertical and horizontal temperature check is used to locate and determine the size of an error in T_1 :

$$\begin{aligned}
&(TH(1,1) \wedge (|\delta T_1^H| \geq |\delta T_2^H|) \wedge \\
&(TV(1,1) \wedge (|\delta T_1^V| \geq |\delta T_2^V|) \wedge \\
&(|\delta T_1^H - \delta T_1^V| \leq \Delta_1^2/B_1^2) \wedge \\
&(TS(1,1) \wedge (|\delta T_1^S - \tau_1|) \leq \Delta_1^2/B_1^2) \ ,
\end{aligned} \tag{99}$$

where

$$\tau_1 = (\delta T_1^H + \delta T_1^V)/2 \tag{100}$$

is the error in T_1.

When T_1 and H_1 and/or T_2 are erroneous the residuals from the vertical and horizontal checks of geopotential are used:

$$\begin{aligned}
&(hH(1,1) \wedge (|\delta H_1^H| \geq |\delta H_2^H|) \wedge \\
&(HV(1,1) \wedge (|\delta H_1^V| \geq |\delta H_2^V|) \wedge \\
&(|\delta H_1^H - \delta H_2^H - \delta H_1^V| \leq |\Delta_1^2|) \ ,
\end{aligned} \tag{101}$$

and

$$\chi_1 = (\delta H_1^H - \delta H_2^H + \delta H_1^V)/2 \tag{102}$$

is used to correct the geopotential height.

When T_1 is erroneous, the following checks are used:

$$\begin{aligned}&(TV(1,1) \wedge (|\delta T_1^V| \geq \\ &|\delta T_2^V|) \wedge TS(1,1) \wedge \\ &(|\delta T_1^V - \delta T_1^S| \leq \Delta_1^2/B_1^2) \wedge \\ &(|\delta T_1^V + \delta T_1^S|/2 \geq 2\cdot\Delta_1^2/B_1^2)\end{aligned} \tag{103}$$

and correction is calculated by

$$\tau = (\delta T_1^V + \delta T_1^S)/2 \tag{104}$$

or the following check

$$\begin{aligned}&TH(2,1) \wedge TS(1,1) \wedge \\ &(|\delta T_1^H - \delta T_1^S|) \leq \Delta_1^2/B_1^2) \wedge \\ &(|\delta T_1^H + \delta T_1^S|/2 \geq 2\cdot\Delta_1^2)\end{aligned} \tag{105}$$

with the correction calculated by

$$\tau = (\delta T_1^H + \delta T_1^S)/2 \ . \tag{106}$$

d. Error τ_i in T_i or error χ_i in H_i or error χ in subsequent geopotential heights starting at H_i for interior levels (i=2,...,n-1).

Error detection analysis of the CQC component residuals is performed for i-th level, if the following condition is fulfilled

$$\begin{gathered}S(2,i-1) \vee S(2,i) \vee Hh(2,i-1) \vee Hh(2,i) \vee \\ TH(2,i) \vee TV(2,i) \vee HH(2,i) \vee HV(2,i)\end{gathered} \tag{107}$$

The procedure starts with a check of the various conditions associated with most common and simplest isolated errors in H_i and T_i.

H_i is considered to be erroneous if the following condition is true

$$S(1,i\text{-}1) \wedge S(1,i) \wedge (|\delta^{i}_{i-1} + \delta^{i+1}_{i}| \le \Delta^{i}_{i-1}) \wedge \\ (|\chi_i| \ge \Delta^{i}_{i-1}) \wedge (|\chi_i - \delta h^{i}_{i-1}| \le \Delta^{i}_{i-1}) \tag{108}$$

where

$$\chi_i = (\delta^{i}_{i-1} - \delta^{i+1}_{i})/2 \tag{109}$$

is the magnitude of the error in H_i.

T_i is considered to be erroneous if the following condition is true

$$S(1,i\text{-}1) \wedge S(1,i) \wedge TH(1,i) \wedge \\ (|\delta^{i}_{i-1}/B_{i-1} - \delta^{i+1}_{i}/B^{i+1}_{i}| \le \Delta^{i+1}_{i}/B^{i}_{i-1}/2) \wedge \\ (|\tau_i - \delta T^{H}_{i}| \le \Delta^{i}_{i-1}/B^{i}_{i-1}/2) \wedge (|\tau_i| \ge \Delta^{i}_{i-1}/B^{i}_{i-1}) \tag{110}$$

where

$$\tau_i = (\delta^{i}_{i-1}/B^{i}_{i-1} + \delta^{i+1}_{i}/B^{i+1}_{i})/2 \ . \tag{111}$$

is the magnitude of the error in T_i.

Then nearly identical conditions are checked using vertical instead of horizontal residuals.

H_i is considered to be erroneous if the following condition is true

$$S(1,i\text{-}1) \wedge S(1,i) \wedge (|\delta^{i}_{i-1} + \delta^{i+1}_{i}| \le \Delta^{i}_{i-1}) \wedge \\ (|\chi_i| \ge \Delta^{i}_{i-1}) \wedge (|\chi_i - \delta H^{V}_{i}| \le \Delta^{i}_{i-1})$$

where

$$\chi_i = (\delta^{i}_{i-1} - \delta^{i+1}_{i})/2 \tag{113}$$

is the magnitude of the error in H_i.

T_i is considered to be erroneous if the following condition is true

$$
\begin{aligned}
&S(1,i-1) \wedge S(1,i) \wedge TV(1,i) \wedge \\
&(|\delta^{i}_{i-1}/B_{i-1} - \delta^{i+1}_{i}/B^{i+1}_{i}| \leq \Delta^{i+1}_{i}/B^{i}_{i-1}/2) \wedge \\
&(|\tau_{i} - \delta T^{V}_{i}| \leq \Delta^{i}_{i-1}/B^{i}_{i-1}/2) \wedge (|\tau_{i}| \geq \Delta^{i}_{i-1}/B^{i}_{i-1})
\end{aligned}
\tag{114}
$$

where

$$
\tau_{i} = (\delta^{i}_{i-1}/B^{i}_{i-1} + \delta^{i+1}_{i}/B^{i+1}_{i})/2 \ . \tag{115}
$$

is the magnitude of the error in T_i.

All values H_j ($j = i,\ldots,n$) are assumed to have an error of magnitude χ (calculated error of thickness) if the following condition is true

$$
\begin{aligned}
&S(1,i-1) \wedge hH(1,i-1) \wedge (S(2,i-1) \vee hH(1,i-1)) \wedge \\
&(|\delta h^{i+1}_{i}| \leq \Delta^{i+1}_{i}/2) \wedge (|\delta h^{i}_{i-1} - \delta^{i}_{i-1}| \leq \Delta^{i}_{i-1}) \wedge \\
&\neg(TH(1,i) \wedge TV(1,i) \wedge TS(1,i))
\end{aligned}
\tag{116}
$$

where $\neg$ is the logical "not" operator and χ is defined by

$$
\chi = \delta^{i}_{i-1} \ . \tag{117}
$$

To detect errors in temperature and geopotential height the following checks are also made.

H_i is considered to be erroneous if the following condition is true

$$
\begin{aligned}
&Hh(1,i) \wedge HV(1,i) \wedge (Hh(2,i) \vee HV(2,i)) \wedge \\
&(|\delta h^{i+1}_{i} - \delta H^{V}_{i}| \leq \Delta^{i}_{i-1}/2)
\end{aligned}
\tag{118}
$$

and magnitude of the error χ_i is defined as

$$
\chi_{i} = (\delta h^{i+1}_{i} + \delta H^{V}_{i})/2 \ ; \tag{119}
$$

T_i is considered to be erroneous, if the following condition is true

$$
\begin{aligned}
&TH(1,i) \wedge TV(1,i) \wedge (TH(2,i) \vee TV(1,i)) \wedge \\
&(|\delta T^{H}_{i} - \delta T^{V}_{i}| \leq \Delta^{i+1}_{i}/B^{i}_{i-1}/2)
\end{aligned}
\tag{120}
$$

and

$$\tau_i = (\delta T_i^H + \delta T_i^V)/2 \tag{121}$$

is the magnitude of the error.

T_i value is considered to be erroneous, if the following condition is true

$$\begin{aligned} &TH(1,i) \wedge TS(1,i) \wedge (TH(2,i) \vee TS(1,i)) \wedge \\ &(|\delta T_i^H - \delta T_i^S| \leq \Delta_i^{i+1}/B_{i-1}^i/2) \end{aligned} \tag{122}$$

and

$$\tau_i = (\delta T_i^H + \delta T_i^S)/2 \tag{123}$$

is the magnitude of the error.

T_i value is considered to be erroneous, if the following condition is true

$$\begin{aligned} &TV(1,i) \wedge TS(1,i) \wedge (TV(2,i) \vee TS(1,i)) \wedge \\ &(|\delta T_i^V - \delta T_i^S| \leq \Delta_i^{i+1}/B_{i-1}^i/2) \end{aligned} \tag{124}$$

where

$$\tau_i = (\delta T_i^V + \delta T_i^S)/2 \tag{125}$$

is the magnitude of the error.

e. Error τ_n in T_n or error χ_n in H_n for interior levels (i=2,...,n-1).

Analysis of the CQC residuals to detect errors at the n-th (top) level is performed, if the following condition is true

$$\begin{aligned} S(2,n-1) \vee Hh(2,n-1) \vee TH(2,n) \vee \\ TV(2,n) \vee HH(2,n) \vee HV(2,n) \end{aligned} \tag{126}$$

The procedure followed is to check for common and simple isolated errors in H_n and T_n using horizontal checks.

H_n value is considered to be erroneous, if the following condition is true

$$S(1,n-1) \wedge hH(1,i-1) \wedge$$
$$(|\chi_n| \geq \Delta_{n-1}^n) \wedge (|\chi_n - \delta h_{n-1}^n| \leq \Delta_{n-1}^n) \tag{127}$$

where

$$\chi_n = \delta_{n-1}^n \tag{128}$$

is the magnitude of the error in H_n.

T_n is considered to be erroneous, if the following condition is true

$$S(1,n-1) \wedge TH(1,n) \wedge$$
$$(|\tau_n - \delta T_n^H| \leq \Delta_{n-1}^n / B_{n-1}^n / 2) \wedge (|\tau_n| \geq \Delta_{n-1}^n / B_{n-1}^n) \tag{129}$$

where

$$\tau_n = \delta_{n-1}^n / B_{n-1}^n \ . \tag{130}$$

is the magnitude of the error in T_n.

Next, almost the same checks are made using vertical checks instead of horizontal checks.

H_n value is considered to be erroneous, if the following condition is true

$$S(1,n-1) \wedge HV(1,n) \wedge$$
$$(|\chi_n| \geq \Delta_{n-1}^n) \wedge (|\chi_n - \delta H_n^V| \leq \Delta_{n-1}^n / 2) \tag{131}$$

where

$$\chi_n = \delta_{n-1}^n \tag{132}$$

is the magnitude of the error in H_n.

T_n value is considered to be erroneous, if the following condition is true

$$S(1,n-1) \wedge TV(1,n) \wedge$$
$$(|\tau_n - \delta T_n^V| \leq \Delta_{n-1}^n / B_{n-1}^n / 2) \wedge (|\tau_n| \geq \Delta_{n-1}^n / B_{n-1}^n) \tag{133}$$

where

$$\tau_n = \delta^n_{n-1}/B^n_{n-1} \ . \tag{134}$$

is the magnitude of the error in T_n.

The following conditions are checked next.

H_n is considered to be erroneous, if the following condition is true

$$\begin{gathered} Hh(1,n) \wedge HV(1,n) \wedge (Hh(2,n) \vee HV(2,n)) \wedge \\ (|\delta h_n^{n+1} - \delta H_n^V| \le \Delta^n_{n-1}/2) \end{gathered} \tag{135}$$

and error χ_n is defined as

$$\chi_n = (\delta h_n^{n+1} + \delta H_n^V)/2 \ ; \tag{136}$$

T_n value is considered to be erroneous, if the condition is true

$$\begin{gathered} TH(1,n) \wedge TV(1,n) \wedge (TH(2,n) \vee TV(1,n)) \wedge \\ (|\delta T_n^H - \delta T_n^V| \le \Delta_1^{n+1}/B^n_{n-1}/2) \end{gathered} \tag{137}$$

where

$$\tau_n = (\delta T_n^H + \delta T_n^V)/2 \tag{138}$$

is the magnitude of the error.

T_n value is considered to be erroneous, if the condition is true

$$\begin{gathered} TH(1,n) \wedge TS(1,n) \wedge (TH(2,n) \vee TS(1,n)) \wedge \\ (|\delta T_n^H - \delta T_n^S| \le \Delta_n^{n+1}/B^n_{n-1}/2) \end{gathered} \tag{139}$$

where

$$\tau_n = (\delta T_n^H + \delta T_n^S)/2 \tag{140}$$

is the magnitude of the error.

T_n value is considered to be erroneous, if the condition is true

$$\begin{gathered} TV(1,n) \wedge TS(1,n) \wedge (TV(2,n) \vee TS(1,n)) \wedge \\ (|\delta T_n^V - \delta T_n^S| \le \Delta_n^{n+1}/B^n_{n-1}/2) \end{gathered} \tag{141}$$

where

$$\tau_n = (\delta T_n^V + \delta T_n^S)/2 \tag{142}$$

is the magnitude of the error.

f. Error in all geopotential heights starting at lowest mandatory level.

Errors in the calculation of the surface pressure results in a constant error χ in the geopotential heights at mandatory levels H_i (i=1,...,n). To identify this error, the following conditions are checked:

$$(\sum_{i=1}^{n-1} S(2,i) = 0) \wedge \neg S(1,1) \wedge \neg S(1,2) \tag{143}$$

When (143) is true the hydrostatic check (strong conditions) does not indicate an error; and

$$\begin{gathered} \neg TH(1,1) \wedge \neg TH(1,2) \wedge \\ (\sum_{i=1}^{n} TH(1,i) < n/3) \wedge (\sum_{i=1}^{n} TH(2,i) = 0) \end{gathered} \tag{144}$$

when (144) is true the horizontal check of temperature (strong checks) does not indicate an error either, and if

$$\begin{gathered} HH(1,1) \wedge (\sum_{i=1}^{n} HH(2,i) \geq 1) \wedge \\ (\sum_{i=2}^{n} (|\delta H_i^H - \delta H_1^H| \leq \Delta_{1_2}) \geq 2n/3) \wedge (|\delta H_2 - \delta H_1^H| \leq \Delta_1^2) \end{gathered} \tag{145}$$

(145) is true then the horizontal check of geopotential heights (weak checks) indicates an error at the first level and a shift of heights for a majority of mandatory levels.

If (143), (144), and (145) are true, all values H_i (i=1,...,n) are corrected by if and only if

$$\chi = \delta H_1^H \ , \tag{146}$$

$$|\chi| \geq 2 \cdot \Delta_1^2 \ . \tag{147}$$

is satisfied.

g. Radiosonde malfunction starting at the lowest or an intermediate level.

If an error has not been resolved by the above tests, the sounding is checked for a radiosonde malfunction.

It is assumed that there are garbled temperatures beginning with the i-th level, resulting in a miscalculation of mandatory geopotential heights, if the following three conditions are true

$$\sum_{j=i}^{n-1} S(2,j) = 0 \tag{148}$$

(148) true means the hydrostatic check does not indicate an error, and

$$TH(1,i) \wedge \left(\sum_{j=i+1}^{n} (\delta T_j^H \cdot \delta T_i^H > 0) \geq \frac{2 \cdot (n-i)}{3} \right) \tag{149}$$

true means the horizontal check of temperature indicates an permanent shift of temperature values for almost all levels above the i-th level, and

$$\begin{gathered} \left(\sum_{j=i}^{n} HH(2,j) \geq 1 \right) \wedge \left(|\delta H_n^H| > \delta H_i^H \right) \wedge \\ \left(\sum_{j=i+1}^{n} (\delta h_{j-1}^{j} \cdot \delta h_{i-1}^{i} > 0) \geq \frac{2 \cdot (n-1)}{3} \right) \end{gathered} \tag{150}$$

true means the horizontal check of geopotential heights indicates an permanent shift of geopotential heights for almost all levels.

When (148), (149), and (150) are true, all temperatures, geopotential heights, humidity, and winds starting with i-th mandatory level, are considered erroneous. No corrections are made to the temperature, geopotential heights, humidity, and winds at these levels.

This concludes the analysis of CQC residuals to identity errors. A feature of the second section of the DMA, where errors in a sounding are identified, is that new error conditions identified during data analysis can be added to the conditions being checked in this section.

If none of the preceding conditions appear to be true, the sounding enters the third section of DMA, where residuals are analyzed for rehabilitation.

In the third section of the DMA, the following conditions are checked for each T_i and H_i $(i=1,\ldots,n)$ for geopotential heights

$$\begin{gathered}((HH(2,i) \wedge HV(2,i)) \vee \\ (HH(1,i) \wedge HV(2,i))) \wedge \\ (\delta H_i^H \cdot \delta H_i^V \geq 0)\end{gathered} \tag{151}$$

and

$$\begin{gathered}((TH(2,i) \wedge TV(2,i)) \vee \\ (TH(1,i) \wedge TV(2,i))) \wedge \\ (\delta T_i^H \cdot \delta T_i^V \geq 0)\end{gathered} \tag{152}$$

for temperatures.

If either (151) or (152) is true, then T_i or H_i is declared to be a suspect value (this means that there is something wrong with this data, but the CQC residuals don't indicate what is wrong).

Hydrostatic residuals identify possible geopotential height errors when the following is true

$$\begin{gathered}((|\delta_{i-1}^i| \geq 1.5 \cdot \Delta_{i-1}^i) \vee \\ (|\delta_i^{i+1}| \geq 1.5 \cdot \Delta_i^{i+1})) \wedge \\ (HH(i,1) \wedge HV(i,1))\end{gathered} \tag{153}$$

and T_i is considered suspect if

$$\begin{aligned}&((|\delta^{i}_{i-1}| \geq 1.5 \cdot \Delta^{i}_{i-1}) \lor \\ &\ (|\delta^{i+1}_{i}| \geq 1.5 \cdot \Delta^{i+1}_{i})) \land \\ &(TH(i,1) \land TV(i,1))\end{aligned} \tag{154}$$

In the DMA, each T_i and H_i value at mandatory levels has a quality control flags assigned to it:

0 - value not checked;

1 - correct value;

2 - suspect value;

3 - erroneous value;

4 - value was erroneous and now is corrected.

Table 10 contains the results of applying the CQC to a global set of upper-air stations. Table 10 show the analysis for temperatures and geopotential heights at mandatory levels.

2. The DMA for winds at mandatory levels

In the current version of the CQC, it is assumed that each error in an upper-air observation is due to garbling of the speed or/and direction values. The residuals of the CQC components for wind are calculated in the terms of the zonal, U, and the meridional components, V. Errors are located by first analyzing the U and V residuals, and then try to determine whether the speed or direction (or both) is the source of the error.

The DMA for wind at mandatory levels consists of three logic sections:

a decision making section for suspicious values;

error identification and error estimation section;

a section for processing the remaining suspicious values.

In the first section, the following checks are carried out if at least one of residual exceeds the corresponding allowable residual:

$$
\begin{aligned}
&|\delta U_i^H| > \Delta U_i^H , \\
&|\delta V_i^H| > \Delta V_i^H , \\
&|\delta U_i^V| > \Delta U_i^V , \\
&|\delta V_i^V| > \Delta V_i^V , \\
&|\delta U_i^G| > \Delta U_i^G , \\
&|\delta V_i^G| > \Delta V_i^G , \\
&|\delta U_i^S| > \Delta U_i^S , \\
&|\delta V_i^S| > \Delta V_i^S , \quad i=1,\dots,n, \\
&|\delta U_{ti}^{i+1}| > \Delta U_{ti}^{i+1} , \\
&|\delta V_{ti}^{i+1}| > \Delta V_{ti}^{i+1} , \quad i=1,\dots,n-1.
\end{aligned}
\tag{155}
$$

where the actual and allowable residual are:

δU_{ti}^{i+1} and ΔU_{ti}^{i+1} - thermal check of U component,

δVt_i^{i+1} and ΔV_{ti}^{i+1} - thermal check of V component,

δU_i^H and ΔU_i^H - horizontal check of U component,

δV_i^H and ΔV_i^H - horizontal check of V component,

δU_i^V and ΔU_i^V - vertical check of U component,

δV_i^V and ΔV_i^V - vertical check of V component,

δU_i^G and ΔU_i^G - geostrophic check of U component,

δV_i^G and ΔV_i^G - geostrophic check of V component,

δU_i^S and ΔU_i^S - U component check using significant levels,

δV_i^S and ΔV_i^S - V component check using significant levels.

If at least one of these inequalities is true, it is assumed that an error is possible in the sounding and the DMA enters the second section, otherwise the sounding is assumed to be correct and the program exits the DMA.

In the second section of the DMA, the logical analysis of residuals and their relations is carried out for all CQC components from the lowest to the highest isobaric surfaces for each sounding.

If an error in U_i or/and V_i (corresponding to an error in speed, S_i, speed or/and the wind direction, A_i, for $i= 1,\dots,n$) is detected and corrected, then U_i or/and V_i and their residuals are recomputed for each CQC component and the sounding again enters the first section of the DMA.

S_i and A_i values are corrected only after an analysis of the magnitudes and signs of the expected corrections, taking into consideration that most errors are usually based upon a mistake or garbling of one digit. Each correction is then rounded off in a manner to ensure that erroneous and corrected values differ by one digit or the sign. If it is not possible to correct the data in this manner, it is assumed that more than one digit is garbled and the value is adjusted to the expected value, rounded to the nearest meter per second for speed and five degree for direction.

The error identification procedures uses a set of logical variables UH, VH, UV, VV, UT, VT, UG, VG, US and VS which are defined as follows

$$
\begin{aligned}
UT(k,i) &= (|\delta U_{ti}^{i+1}| > k \cdot \Delta U_{ti}^{i+1}/2)\ , \\
VT(k,i) &= (|\delta V_{ti}^{i+1}| > k \cdot \Delta V_{ti}^{i+1}/2)\ , \quad i=1,\dots,n-1, \\
UH(k,i) &= (|\delta U_i^H| > k \cdot \Delta U_i^H/2)\ , \\
VH(k,i) &= (|\delta V_i^H| > k \cdot \Delta V_i^H/2)\ , \\
UV(k,i) &= (|\delta U_i^V| > k \cdot \Delta U_i^V/2)\ , \\
VV(k,i) &= (|\delta V_i^V| > k \cdot \Delta V_i^V/2)\ , \\
US(k,i) &= (|\delta U_i^S| > k \cdot \Delta U_i^S/2)\ , \\
VS(k,i) &= (|\delta V_i^S| > k \cdot \Delta V_i^S/2)\ , \\
UG(k,i) &= (|\delta U_i^S| > k \cdot \Delta U_i^G/2)\ , \\
VG(k,i) &= (|\delta V_i^G| > k \cdot \Delta V_i^G/2)\ , \quad i=1,\dots,n;\ k = 1,2.
\end{aligned}
\tag{156}
$$

If any of these condition is true, than the variable will have a value of one. When the condition is false the variable is set to zero.

If a test in (156) is satisfied (true) for k = 1, the error is classified as weak and strong for k = 2.

With this notation, a decision concerning the existence of an error is made according to the following rules and in the following order:

If the following condition is satisfied, an error analysis of the CQC residuals at the i-th level (i=1,...,n) is performed

$$\begin{aligned} &UH(2,i) \vee VH(2,i) \vee \\ &UV(2,i) \vee VV(2,i) \vee \\ &US(2,i) \vee VS(2,i) \end{aligned} \tag{157}$$

It is assumed that the residuals δU_{is} and δV_{is} give the error when the following condition

$$\begin{aligned} &(US(1,i) \vee VS(1,i)) \wedge \\ &(|\delta U_i^H - \delta U_i^S| \leq |\delta U_i^H|) \wedge (|\delta V_i^H - \delta V_i^S| \leq |\delta V_i^H|) \wedge \\ &(|\delta U_i^V - \delta U_i^S| \leq |\delta U_i^V|) \wedge (|\delta V_i^V - \delta V_i^S| \leq |\delta V_i^V|) \end{aligned} \tag{158}$$

is true.

The next task is to transform the residuals from U and V components to speed S and direction A. The conversion procedure is shown in Figs. 49 and 50. The idea is that a allowable region ABCD for the U and V wind components is defined by the allowable residuals ΔU_{is} *and* ΔV_{is}. The transformed wind, wind speed and direction, must lie in a limiting region A'B'C'D'. Using δU_{is} and δV_{is} the errors in wind speed and direction at the i-th mandatory level (E_{si} and/or E_{ai}) are found. Corrected values of S and A can be found in A'B'C'D' using the condition that minimal corrections are made (for example, correct only wind speed or direction).

If the procedure outlined in the previous paragraph is successful it is assumed that the wind speed at the i-th level has error E_{si} and/or the wind direction has error E_{ai}. Otherwise, the following checked is performed

$$(UH(1,i) \vee VH(1,i)) \wedge (UV(1,i) \vee (VV(1,i)) \qquad (159)$$

and error in wind speed and directions is determined by using the following error estimates

$$\begin{aligned} U_{ei} &= (\delta U_i^H + \delta U_i^V)/2 \ , \\ V_{ei} &= (\delta V_i^H + \delta V_i^V)/2 \ . \end{aligned} \qquad (160)$$

Here the conversion procedure of errors U_{ci} and V_{ci} to errors E_{si} and/or E_{ai} is repeated (see Figs. 49 and 50) with the difference that the primary allowable region ABCD is used for both limiting regions. The first region is defined by the allowable residuals of the horizontal check (ΔU_{iH} and ΔV_{iH}) and the second is defined by the allowable residuals of the vertical check (ΔU_{iV} and ΔV_{iV}).

This ends the CQC residual analysis to identity possible errors in the mandatory level wind data. The powerful feature of this second section of the DMA, where sounding errors are identified, is that new error checks identified during data analysis, can be added to the DMA.

If none of the preceding tests are true, the sounding enters the third section of DMA, where residuals are analyzed for restoration.

In the third section of DMA, the following tests are conducted on each U_i and V_i ($i=1,\ldots,n$) value:

$$\begin{aligned} &(|\delta U_i^H| > 2.0 \cdot \Delta U_i^H \ , \\ &(|\delta V_i^H| > 2.0 \cdot \Delta V_i^H \ , \\ &(|\delta U_i^V| > 1.5 \cdot \Delta U_i^V \ , \\ &(|\delta V_i^V| > 1.5 \cdot \Delta V_i^V \ , \quad i=1,\ldots,n, \end{aligned} \qquad (161)$$

where

δUi^H and ΔU_i^H - residuals of the horizontal check of U component,

δV_i^H and ΔV_i^H - residuals of the horizontal check of V component,

δU_i^V and ΔU_i^V - residuals of the vertical check of U component,

δV_i^V and ΔV_i^V - residuals of the vertical check of V component.

If any of the conditions in (161) are true, then both wind speed and directions at the mandatory level are flagged as suspicious.

After processing the sounding via the CQC procedures for winds, all S_i and A_i at mandatory levels have quality control flags assigned (it should be noted that CQC does not currently correct any wind data):

0 - value not checked;

1 - correct value;

2 - suspect value;

3 - erroneous value.

Table 11 shows the results from the CQC wind speed and direction analysis at mandatory levels for a global set of upper-air stations.

3. DMA for humidity at mandatory levels

Research into the nature of potential errors in humidity observations is needed. In the current version of CQC, it is assumed that each error in an upper-air observation is due to a garbling of the numbers.

The DMA for humidity data at mandatory levels consists of three sections:

a decision making section for suspicious values;

a error identification and error estimation section,

a section to process the remaining suspicious values.

In the first section, potential errors are identified when at least one of the residuals exceeds the corresponding allowable residual:

$$
\begin{aligned}
|\delta D_i^H| &> \Delta D_i^H , \\
|\delta D_i^V| &> \Delta D_i^V , \\
|\delta D_i^S| &> \Delta D_i^S , \quad i=1,\ldots,n,
\end{aligned}
\tag{162}
$$

where

δD_i^H and ΔD_i^H - actual and allowable residuals for the horizontal humidity check,

δD_i^V and ΔD_i^V - actual and allowable residuals of the vertical humidity check, and

δD_i^S and ΔD_i^S - actual and allowable residuals of the significant level humidity check.

If at least one of these inequalities is true, it is assumed that an error is possible in the sounding and it enters the second section of the DMA, otherwise the sounding is assumed to be correct and processing in the DMA is terminated.

In the second section of the DMA, a logical analysis of the residuals is performed from the lowest to the highest pressure levels in each sounding.

The error identification procedures use a set of logical variables DH, DV and DS which are defined as

$$
\begin{aligned}
DH(k,i) &= (|\delta D_i^H| > k \cdot \Delta D_i^H/2) , \\
DV(k,i) &= (|\delta D_i^V| > k \cdot \Delta D_i^V/2) , \\
DS(k,i) &= (|\delta D_i^S| > k \cdot \Delta D_i^S/2) , \quad i=1,\ldots,n;\ k = 1,2.
\end{aligned}
\tag{163}
$$

When the condition is true the value is set to one and zero when false.

If an error is detected, a correction is estimated. The corrected value is given by "D-d" in the equations below. Then the logical variables are recomputed using the correction and the sounding again enters the first section of the DMA.

If the conditions in (163) are true for k = 1, the error classification is called weak and strong for k = 2.

With this notation, the existence of an error is determined according to the following rules and in the following order:

Analysis of CQC component residuals for error detection in the values at i-th level (i=1,...,n) is carried out if the following condition is true

$$DH(2,i) \vee DH(2,i) \vee DS(2,i) \tag{164}$$

Then checks are made to determine the type of error

$$\begin{aligned} &DS(1,i) \wedge (|\delta d| \geq 7°C) \wedge \\ &DV(1,i) \wedge (\delta D_i^H \times d > 0) \wedge \\ &DH(1,i) \wedge (\delta D_i^V \times d > 0) \end{aligned} \tag{165}$$

where

$$d = \delta D_i^S \ . \tag{166}$$

If (165) is true the corrected dewpoint depression is

D_i-d.

The relative humidity R_i is calculated from the corrected dewpoint depression D_i-d and must be in the range

$$0 \leq R_i \leq 100 \ \% \ , \tag{167}$$

If all the above conditions are true, then it is assumed that observed value D_i is erroneous.

The next error check is

$$\begin{aligned} UH(1,i) \wedge UV(1,i) \wedge (|d| \geq 7°C) \wedge \\ (|\delta D_i^H - \delta D_i^V| \leq 5°C) \end{aligned} \tag{168}$$

where

$$d = (\delta D_i^H + \delta D_i^V)/2 \ , . \tag{169}$$

The corrected value D-d must satisfy (167).

If there is evidence from the horizontal or vertical checks that a garbling of humidity data at neighboring stations or adjacent levels, then significant levels can be used to check the data

$$DH(1,i) \wedge DS(1,i) \wedge (|d| \geq 7°C) \wedge (\delta D_i^H \cdot d \geq 0) \tag{170}$$

or

$$DV(1,i) \wedge DS(1,i) \wedge (|d| \geq 7°C) \wedge (\delta D_i^V \cdot d \geq 0) \tag{171}$$

where

$$d = \delta D_i^S. \tag{172}$$

Once again the corrected value D-d must satisfy (167).

This is the end of the analysis of the CQC residuals to identity errors in the humidity data at mandatory levels. A feature of this second section of the DMA, where errors are identified, is that new error conditions identified during data analysis can be added to the conditions being checked in this section.

If the tests (165), (168), (170), and (171) are false, the sounding enters the third section of the DMA.

In the third section of DMA, the following conditions are checked for $i=1,\dots,n$:

$$\begin{gathered} 2212(|\delta D_i^H| > 2.0 \cdot \Delta D_i^H , \\ (|\delta D_i^V| > 1.5 \cdot \Delta D_i^V , \\ (R_i < 0\%) \vee (R_i > 100\%) \ , \quad i=1,\dots,n, \end{gathered} \tag{173}$$

where

δD_{iH} and ΔD_{iH} - actual and allowable residuals from the horizontal check of humidity,

δD_{iV} and ΔD_{iV} - actual and allowable residuals from the vertical check of humidity,

R_i- the relative humidity calculated from the dewpoint depression, D_i.

If any of these conditions in (173) is true, then the dewpoint depression is considered to be suspicious. However, a correction cannot be made. The datum is flagged and the program exits the DMA.

After CQC processes each sounding, the dewpoint depression, D_i, at mandatory levels has a quality control flag assigned (it should be noted that the current version of CQC does not correct humidity data):

0 - value was not checked;

1 - correct value;

2 - suspicious value;

3 - erroneous value.

Table 12 shows CQC results from dewpoint depression data at mandatory levels for a global set of upper-air stations. The current version of CQC does not check humidity data above 300 hPa. Climatic data does not exist for these levels. Some data are unchecked below 300 hPa, this is due to the fact that the temperature data at these levels are erroneous or suspect. In these cases, the humidity value can not be checked.

4. The DMA for all variables at significant levels

It is assumed that errors in geopotential height, temperature, U and V wind component and humidity at significant levels are due to garbling of the data.

The DMA for the meteorological variables at significant levels consists of three sections:

a decision making section to locate suspicious data;

an error identification and correction section;

a section to process the remaining suspicious data.

In the first section, potential errors are identified when the residual exceeds the allowable residual for that parameter:

$$
\begin{aligned}
|\delta F_i^M| &> \Delta F_i^M , \\
|\delta F_i^D| &> \Delta F_i^D , \\
|\delta F_i^U| &> \Delta F_i^U , \quad i=1,\ldots,n,
\end{aligned}
\tag{174}
$$

where n is the number of significant levels and F represents the temperature, geopotential height, humidity, and the U and V components of the wind;

δF_i^M and ΔF_i^M - the actual and allowable residuals from linear interpolation of F from adjacent mandatory levels,

δF_i^D and ΔF_i^D - the actual and allowable residuals from linear extrapolation of F from higher significant and/or mandatory levels,

δF_{iU} and ΔF_i^U - the actual and allowable residuals from the linear extrapolation of F from lower significant and/or mandatory levels.

If at least one of these inequalities is true, it is assumed that an error is possible in the sounding and the DMA enters the second section, otherwise the sounding is assumed to be correct and processing in the DMA is terminated.

In the second section of the DMA, a logical analysis of the residuals is performed from the lowest to the highest pressure levels in each sounding.

If an error in F_i is detected, a correction is made (F_i - d) and then F_i and the residuals are recomputed and the sounding again enters the first section of the DMA.

The error identification procedures uses a set of logical variables FM, FD and FU which are defined as

$$
\begin{aligned}
FM(k,i) &= (|\delta F_i^M| > k \cdot \Delta F_i^M/2) , \\
FD(k,i) &= (|\delta F_i^D| > k \cdot \Delta F_i^D/2) , \\
FU(k,i) &= (|\delta F_i^U| > k \cdot \Delta F_i^U/2) , \quad i=1,\ldots,n;\ k = 1,2.
\end{aligned}
\tag{175}
$$

If any of these condition is true, then the variable will have a value of one. When the condition is false the variable is set to zero.

If a test in (175) is true for k = 1, the error classified as weak and strong for k = 2.

With this notation, a decision concerning the existence of an error is made according to the following rules and in the following order:

The first test made is to determine at the i-th level (i=1,...,n) if the following simple condition is true

$$FM(1,i) \tag{176}$$

If (176) is true then the following two checks are made

$$\begin{gathered} FM(2,i) \wedge \\ FD(2,i) \wedge (\delta F_i^D \cdot d > 0) \end{gathered} \tag{177}$$

or

$$\begin{gathered} FM(2,i) \wedge \\ FU(2,i) \wedge (\delta F_i^U \cdot d > 0) \end{gathered} \tag{178}$$

where

$$d = \delta F_i^M . \tag{179}$$

If (177) or (178) is true the error in the data has magnitude d.

The next error test is

$$\begin{gathered} FM(1,i) \wedge FD(1,i) \wedge (FM(2,i) \\ \vee FD(2,i)) \wedge (|\delta F_i^M - \delta F_i^D| \le \Delta F_i^M) \end{gathered} \tag{180}$$

or

$$FM(1,i) \wedge FU(1,i) \wedge$$
$$(FM(2,i) \vee FU(2,i)) \wedge \tag{181}$$
$$(|\delta F_i^M - \delta F_i^U| \leq \Delta F_i^M)$$

and the error is assumed to be

$$d = \delta F_i^M \ .$$

The last error test is

$$FU(2,i) \wedge FD(2,i) \wedge (\delta F_i^D \cdot \delta F_i^U > 0) \tag{183}$$

and error in this case is

$$d = (\delta F_i^U + \delta F_i^V)/2 \ . \tag{184}$$

This is the end of the analysis of the CQC residuals to identity a possible error in observations at significant levels. A feature of this second section of the DMA, where sounding errors are identified, is that new error conditions, identified during data analysis can be added to the conditions being checked in this section.

If (176) is true and (177), (178), (180), (181), and (183) are false the sounding enters the third section of the DMA.

In the third section of the DMA, the following condition is checked for each variable F_i:

$$FM(2,i) \wedge (FD(1,i) \vee FU(1,i)) \ . \tag{185}$$

If this condition is true, then F_i at the significant level is considered to be suspicious. A correction is not made.

After CQC processes each sounding, F_i (geopotential height, temperature, wind, humidity) at each significant level has a quality control flags assigned to it (it should be noted that the current version of CQC does not correct data at significant levels):

0 - value not checked;

1 - correct value;

2 - suspect value;

3 - erroneous value.

The CQC processes the wind in component form. Flags are assigned in component form. The flags must be converted to a wind speed and direction format. The conversion is a very simple procedure.

First, if a U or V wind component has quality flag 2 assigned (suspicious value) then both the speed and direction will have this flag value assigned. Second, if a U or V wind component has quality flag 3 assigned (erroneous value) then both the speed and direction have this flag value assigned.

Tables 13, 14, and 15 show the results of a CQC analysis for geopotential heights, temperature, wind, and dewpoint depression at significant levels for a global set of upper-air stations.

IV. Acknowledgements.

The work of Oleg A. Alduchov was supported by the Resident Research Associateship of the National Research Council of the National Academy of Sciences and funded by the NCDC. The Comprehensive Aerological Research Data Set (CARDS) program is supported by the U. S. Department of Energy under contract No. DE-AI05-90ER61011. The CARDS program is also supported by the Climate and Global Change program of NOAA and the National Climatic Data Center. We thank Drs. William G. Collins and Trevor Wallis for reviewing this paper.

V. References

Alduchov, O. A., 1982: Combined quality control of height and temperature for isobaric surfaces in FGGE upper-air reports. GARP International Conf. on the Scientific Results of the Monsoon Experiments. Extended Abstract and Panel Session. Denpasar, Bali, Global Atmospheric Research Program, WMO, 8.15-8.18.

Alduchov O.A., 1983: The complex control of FGGE Level-II upper-air data. Meteorology and Hydrology, 1983, N 12.

Alduchov, O. A., and R. G. Reitenbach, 1990: Horizontal geopotential and temperature correlations in the free air. Meteor. Gidrol., 8, 76-83 (in Russian).

Antsipovich, V. A., 1980: Complex quality control of height and temperature at mandatory isobaric surfaces. Proceedings of USSR Hydrometeorological Center, No. 217, (in Russian).

Gandin, L. S., 1963: Objective analysis of meteorological fields. Hydrometeoisdat, 287 pp., (in Russian).

_____, 1969: On automatic quality control of current meteorological information. Meteor. Gidrol., No. 3, 3-13, (in Russian).

_____, and R. L. Kagan, 1976: Statistical methods of meteorological data interpretation. Hydrometeoisdat, 360pp., (in Russian).

_____, 1988: Complex Quality Control of Meteorological Observations. Mon. Wea. Rev. 116, 1137-1156.

Hawson, C. L., 1970: Performance requirements of aerological instruments. WMO Technical Note No. 112, WMO No. 267, 49pp.

Hooper, A. H., 1975: Upper-air sounding studies. Vol. 1, Studies on radiosonde performance. WMO Technical Note No. 140, WMO No. 394, 1-110.

Liberman J. M., 1980: An algorithm for the complex hydrostatic-temporal quality control of upper-air information. Main Geophysical Observatory, Trudy, No. 431, 23-30 (in Russian).

Methodical Guidelines for hydrometeorological stations and sites 1973: No. 4, Part III a - L., Hydrometeoisdat, 256. (in Russian).

Parfiniewicz, J., 1976: Complex quality control of upper-air information. Methodical Guidelines. USSR Hydrometeorological Center, 66 pp., (in Russian).

Reitenbach, R. G., and A. M. Sterin, 1987: Creation of climatic gridded datasets by objective analysis, Proceedings of AURIHMI-WDC, 135, 1-45 (in Russian).

Table 1. Response of the CQC components to different types of gross errors at mandatory levels. The first character in the type of error is: H horizontal, V vertical, G geostrophic, T thermal, S using significant levels. The second character represents: H geopotential height; h thickness; T temperature; W wind, R humidity. "**" represents the hydrostatic test. In the table '+' indicates the presence of a response, '±' the presence of a weak response, and '-' the absent of a response.

Type of error	**	HH	Hh	VH	HT	VT	ST	HW	VW	GW	TW	SW	HR	VR	TR	SR
1. Error in station location	-	+	+	±	+	±	-	+	±	+	+	-	+	±	-	-
2. Observation error in temperature	-	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-
3. Computation error in geopotential	+	±	+	+	-	-	-	-	-	-	-	-	-	-	-	-
4. Distortion of single H-value	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-
5. Distortion of single T-value	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-
6. Distortion of single value of wind speed or direction	-	-	-	-	-	-	-	+	+	+	+	+	-	-	-	-
7. Distortion of single value of humidity	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+

Table 2. Mean and rms values of the hydrostatic residuals calculated from a world-wide dataset for 0 UTC, 01/15/89. N is the number of observations, ϕ is the latitude, δ the mean geopotential residual value, and E is the rms value of the residuals.

LAYER	$-30° < \phi \le 30°$			$30° \le \phi < 60°$			$60° \le \phi$ 90°		
P hpa	N	δ	E	N	δ	E	N	δ	E
20- 10	23	4.2	20.2	68	-4.1	39.8	11	-3.1	29.7
30- 20	66	0.4	11.7	202	4.1	16.1	35	1.6	9.7
50- 30	80	7.5	15.5	242	1.0	14.5	47	0.1	13.0
70- 50	93	-0.7	11.8	291	1.6	10.7	60	0.2	6.6
100- 70	104	-9.1	19.1	317	-1.5	10.5	73	1.4	5.8
150- 100	123	-5.7	13.6	361	2.8	11.9	91	4.9	10.4
200- 150	139	0.6	5.9	370	2.2	9.5	97	1.9	7.7
250- 200	148	1.0	5.2	377	-0.1	6.5	102	-0.7	6.6
300- 250	145	1.4	4.7	389	-0.9	5.6	103	-0.2	5.6
400- 300	146	3.1	9.0	402	-0.7	6.7	102	-0.8	7.0
500- 400	145	2.4	6.5	403	2.3	6.4	100	0.4	5.4
700- 500	141	6.2	10.8	399	5.0	10.6	100	4.2	9.0
850- 700	134	7.2	9.2	388	4.1	7.2	102	4.3	6.8
1000- 850	112	7.3	11.4	195	2.6	7.1	40	7.6	11.0

Table 3. Mean and rms values of the hydrostatic temperature residuals calculated from a global dataset for 0 UTC, 01/15/89. N is the number of observations, ϕ is latitude, $\bar{\alpha}$ is the mean normalized temperature residuals, and β rms value of the normalized temperature residuals.

LAYER	$-30° < \phi \leq 30°$			$30° \leq \phi < 60°$			$60 \leq \phi$ $90°$		
P hpa	N	$\bar{\alpha}$	β	N	$\bar{\alpha}$	β	N	$\bar{\alpha}$	β
20- 10	23	0.4	2.0	68	-0.1	4.0	11	-0.3	2.9
30- 20	66	0.1	1.9	202	0.5	2.2	35	0.2	1.6
50- 30	80	1.0	2.0	242	0.0	1.9	47	0.0	1.7
70- 50	93	-0.1	2.4	291	0.3	2.2	60	0.0	1.3
100- 70	104	-1.9	3.4	317	-0.2	1.9	73	0.2	1.1
150- 100	123	-0.8	2.2	361	0.4	1.9	91	0.8	1.7
200- 150	139	0.2	1.4	370	0.5	2.2	97	0.4	1.8
250- 200	148	0.2	1.5	377	-0.1	2.0	102	-0.2	2.0
300- 250	145	0.5	1.7	389	-0.3	2.0	103	0.0	2.1
400- 300	146	0.6	2.1	402	-0.1	1.6	102	-0.1	1.6
500- 400	145	0.6	1.9	403	-0.7	1.9	100	0.1	1.6
700- 500	141	1.3	2.2	399	0.9	2.0	100	0.8	1.8
850- 700	134	2.6	3.3	388	1.4	2.5	102	1.5	2.4
1000- 850	112	2.6	4.2	195	1.1	3.0	40	3.2	4.6

Table 4. Vertical correlations of geopotential between mandatory levels.

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
latitude < -60 °															
1000	100	93	86	71	63	57	56	54	50	44	32	26	18	12	0
850	93	100	96	82	74	67	65	63	59	50	37	28	19	15	8
700	86	96	100	92	85	79	77	75	70	60	45	36	25	20	15
500	71	82	92	100	98	94	92	88	81	68	52	44	31	27	25
400	63	74	85	98	100	98	95	91	84	71	57	50	31	29	25
300	57	67	79	94	98	100	98	94	87	74	61	55	37	33	27
250	56	65	77	92	95	98	100	97	92	80	66	57	44	41	31
200	54	63	75	88	91	94	97	100	98	91	81	73	55	50	39
150	50	59	70	81	84	87	92	98	100	97	89	83	67	61	48
100	44	50	60	68	71	74	80	91	97	100	97	92	81	74	57
70	32	37	45	52	57	61	66	81	89	97	100	98	91	83	63
50	26	28	36	44	50	55	57	73	83	92	98	100	96	90	72
30	18	19	25	31	31	37	44	55	67	81	91	96	100	97	77
20	12	15	20	27	29	33	41	50	61	74	83	90	97	100	86
10	0	8	15	25	25	27	31	39	48	57	63	72	77	86	100
-60° < latitude < -30°															
1000	100	62	46	38	31	27	23	17	10	6	0	-4	-6	-11	0
850	62	100	90	75	67	60	57	49	39	35	34	26	16	18	0
700	46	90	100	93	87	80	75	69	59	53	45	33	22	23	0
500	38	75	93	100	98	93	88	84	75	67	56	41	29	26	0
400	31	67	87	98	100	98	94	90	82	73	61	45	32	29	0
300	27	60	80	93	98	100	98	95	87	77	65	49	36	32	0
250	23	57	75	88	94	98	100	98	91	81	68	52	41	35	0
200	17	49	69	84	90	95	98	100	97	87	72	56	44	38	0
150	10	39	59	75	82	87	91	97	100	95	79	64	52	43	0
100	6	35	53	67	73	77	81	87	95	100	91	79	67	57	0
70	0	34	45	56	61	65	68	72	79	91	100	94	79	62	0
50	-4	26	33	41	45	49	52	56	64	79	94	100	92	76	0
30	-6	16	22	29	32	36	41	44	52	67	79	92	100	94	0
20	-11	18	23	26	29	32	35	38	43	57	62	76	94	100	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100

Table 4. Continued

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
							-30° < latitude < 30°								
1000	100	76	54	34	24	15	9	5	0	-2	0	1	3	0	7
850	76	100	87	62	47	31	23	17	12	15	16	14	14	12	14
700	54	87	100	85	70	52	43	36	31	33	33	28	27	25	22
500	34	62	85	100	94	81	73	67	60	56	50	44	39	36	35
400	24	47	70	94	100	93	87	81	74	65	56	47	41	39	37
300	15	31	52	81	93	100	97	94	87	75	63	53	47	45	42
250	9	23	43	73	87	97	100	97	92	79	66	56	48	45	42
200	5	17	36	67	81	94	97	100	96	83	68	58	49	46	41
150	0	12	31	60	74	87	92	96	100	92	76	64	52	48	41
100	-2	15	33	56	65	75	79	83	92	100	90	77	60	54	44
70	0	16	33	50	56	63	66	68	76	90	100	92	75	67	54
50	1	14	28	44	47	53	56	58	64	77	92	100	91	83	68
30	3	14	27	39	41	47	48	49	52	60	75	91	100	96	82
20	0	12	25	36	39	45	45	46	48	54	67	83	96	100	91
10	7	14	22	35	37	42	42	41	41	44	54	68	82	91	100
							30° < latitudes < 60°								
1000	100	87	71	52	45	42	41	40	39	35	28	22	16	13	17
850	87	100	91	72	65	61	59	58	55	49	40	31	21	15	14
700	71	91	100	92	86	82	80	78	72	62	50	38	25	18	13
500	52	72	92	100	98	96	93	90	82	69	54	41	28	20	12
400	45	65	86	98	100	98	96	92	83	70	54	41	27	19	11
300	42	61	82	96	98	100	99	96	87	73	57	43	29	21	12
250	41	59	80	93	96	99	100	98	91	77	61	47	32	23	14
200	40	58	78	90	92	96	98	100	96	85	70	55	39	29	19
150	39	55	72	82	83	87	91	96	100	95	83	70	52	41	28
100	35	49	62	69	70	73	77	85	95	100	95	85	68	57	40
70	28	40	50	54	54	57	61	70	83	95	100	96	84	73	55
50	22	31	38	41	41	43	47	55	70	85	96	100	94	86	68
30	16	21	25	28	27	29	32	39	52	68	84	94	100	97	84
20	13	15	18	20	19	21	23	29	41	57	73	86	97	100	93
10	17	14	13	12	11	12	14	19	28	40	55	68	84	93	10

Table 4. Continued

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
					latitude > 60°										
1000	100	87	76	59	52	47	45	41	36	29	26	25	27	29	26
850	87	100	95	82	75	70	68	64	57	48	41	37	34	32	23
700	76	95	100	95	90	86	83	79	72	60	52	45	39	35	22
500	59	82	95	100	99	97	95	90	82	70	60	51	41	34	19
400	52	75	90	99	100	99	97	93	85	72	62	53	42	33	17
300	47	70	86	97	99	100	99	96	89	77	66	56	45	35	18
250	45	68	83	95	97	99	100	98	92	81	72	62	50	40	22
200	41	64	79	90	93	96	98	100	97	90	81	72	60	50	30
150	36	57	72	82	85	89	92	97	100	97	91	84	73	63	42
100	29	48	60	70	72	77	81	90	97	100	98	94	85	76	56
70	26	41	52	60	62	66	72	81	91	98	100	98	92	84	67
50	25	37	45	51	53	56	62	72	84	94	98	100	97	92	77
30	27	34	39	41	42	45	50	60	73	85	92	97	100	98	89
20	29	32	35	34	33	35	40	50	63	76	84	92	98	100	95
10	26	23	22	19	17	18	22	30	42	56	67	77	89	95	100

Table 5. Vertical correlation of temperature between mandatory levels.

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
							latitude < -60°								
1000	100	40	30	25	20	15	10	8	6	4	3	2	1	0	-
850	40	100	66	41	38	9	-16	-20	-15	-26	-25	-17	-5	-2	-
700	30	66	100	72	60	9	-39	-39	-38	-39	-35	-30	-14	1	-
500	25	41	72	100	86	7	-58	-50	-47	-48	-39	-31	-17	8	-
400	20	38	60	86	100	21	-52	-47	-47	-45	-33	-25	-3	4	-
300	15	9	9	7	21	100	30	17	2	-6	-3	-2	9	1	-
250	10	-16	-39	-58	-52	30	100	81	64	51	42	29	8	-3	-
200	8	-20	-39	-50	-47	17	81	100	79	62	44	31	5	-2	-
150	6	-15	-38	-47	-47	2	64	79	100	79	60	39	14	0	-
100	4	-26	-39	-48	-45	-6	51	62	79	100	81	61	31	11	-
70	3	-25	-35	-39	-33	-3	42	44	60	81	100	77	50	31	-
50	2	-17	-30	-31	-25	-2	29	31	39	61	77	100	68	49	-
30	1	-5	-14	-17	-3	9	8	5	14	31	50	68	100	70	-
20	0	-2	1	8	4	1	-3	-2	0	11	31	49	70	100	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100
							-60° < latitude < -30°								
1000	100	42	29	26	32	32	13	1	-11	-24	-16	-23	-10	-22	-
850	42	100	68	46	43	34	18	-4	-29	-45	-45	-27	-3	-11	-
700	29	68	100	68	60	43	18	-9	-39	-57	-47	-29	-3	-23	-
500	26	46	63	100	83	62	25	-9	-47	-58	-47	-30	-8	-13	-
400	32	43	60	83	100	78	33	-7	-47	-57	-45	-33	-17	-22	-
300	32	34	43	62	78	100	61	9	-38	-51	-37	-31	-10	-17	-
250	13	18	18	25	33	61	100	61	-1	-32	-26	-17	-4	-1	-
200	1	-4	-9	-9	-7	9	61	100	45	-7	-11	-1	7	-1	-
150	-11	-29	-39	-47	-47	-38	-1	45	100	44	28	17	-5	2	-
100	-24	-45	-57	-58	-57	-51	-32	-7	44	100	62	45	5	14	-
70	-16	-45	-47	-47	-45	-37	-26	-11	28	62	100	59	27	24	-
50	-23	-27	-29	-30	-33	-31	-17	-1	17	45	59	100	48	46	-
30	-10	-3	-3	-8	-17	-10	-4	7	-5	5	27	48	100	47	-
20	-22	-11	-23	-13	-22	-17	-1	-1	2	14	24	46	47	100	-
10	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100

Table 5. Continued

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
-30° < latitude < 30°															
1000	100	35	26	6	0	1	5	8	12	10	-6	0	4	7	-4
850	35	100	42	18	14	6	9	5	11	6	-5	-1	7	10	4
700	26	42	100	38	33	24	17	14	3	-8	-7	4	9	10	-1
500	6	18	38	100	65	44	33	21	-8	-28	-14	1	6	15	9
400	0	14	33	65	100	67	43	31	-15	-33	-16	4	7	13	7
300	1	6	24	44	67	100	65	53	-10	-30	-14	9	4	10	4
250	5	9	17	33	43	65	100	61	15	-22	-13	-8	-7	1	-1
200	8	5	14	21	31	53	61	100	27	-13	-4	10	8	10	1
150	12	11	3	-8	-15	-10	15	27	100	34	13	-7	-4	-2	-12
100	10	6	-8	-28	-33	-30	-22	-13	34	100	25	0	-2	1	6
70	-6	-5	-7	-14	-16	-14	-13	-4	13	25	100	40	25	24	3
50	0	-1	4	1	4	9	-8	10	-7	0	40	100	53	39	0
30	4	7	9	6	7	4	-7	8	-4	-2	25	53	100	56	13
20	7	10	10	15	13	10	1	10	-2	1	24	39	56	100	35
10	-4	4	-1	9	7	4	-1	1	-12	6	3	0	13	35	100
30° < latitude < 60°															
1000	100	61	52	44	37	20	-11	-22	-22	-18	-13	-6	0	0	-7
850	61	100	79	63	53	20	-19	-37	-31	-26	-19	-11	-4	-3	-3
700	52	79	100	81	68	23	-28	-49	-42	-36	-25	-14	-6	-5	-4
500	44	63	81	100	88	36	-28	-56	-48	-39	-28	-15	-4	-1	-7
400	37	53	68	88	100	54	-18	-51	-46	-40	-28	-14	-4	-1	-6
300	20	20	23	36	54	100	46	-1	-8	-15	-11	-3	-1	-3	-9
250	-11	-19	-28	-28	-18	46	100	63	43	22	16	9	4	0	-3
200	-22	-37	-49	-56	-51	-1	63	100	77	50	35	24	11	2	-1
150	-22	-31	-42	-48	-46	-8	43	77	100	69	54	39	19	6	-4
100	-18	-26	-36	-39	-40	-15	22	50	69	100	76	59	37	17	-3
70	-13	-19	-25	-28	-28	-11	16	35	54	76	100	77	57	37	11
50	-6	-11	-14	-15	-14	-3	9	24	39	59	77	100	74	54	17
30	0	-4	-6	-4	-4	-1	4	11	19	37	57	74	100	79	40
20	0	-3	-5	-1	-1	-3	0	2	6	17	37	54	79	100	65
10	-7	-3	-4	-7	-6	-9	-3	-1	-4	-3	11	17	40	65	100

Table 5. Continued

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
					latitude > 60°										
1000	100	67	60	52	45	25	0	-2	6	11	6	2	0	-3	-11
850	67	100	84	70	59	22	-11	-13	1	11	4	0	-5	-10	-20
700	60	84	100	86	74	28	-16	-20	-2	8	2	-2	-8	-14	-25
500	52	70	86	100	90	39	-15	-22	-2	9	2	-2	-10	-20	-30
400	45	59	74	90	100	57	-3	-15	3	13	4	0	-9	-21	-33
300	25	22	28	39	57	100	62	36	42	42	32	22	8	-7	-23
250	0	-11	-16	-15	-3	62	100	83	73	61	53	43	28	12	3
200	-2	-13	-20	-22	-15	36	83	100	88	74	66	56	39	23	10
150	6	1	-2	-2	3	42	73	88	100	89	82	72	51	30	10
100	11	11	8	9	13	42	61	74	89	100	92	84	67	45	17
70	6	4	2	2	4	32	53	66	82	92	100	94	81	63	36
50	2	0	-2	-2	0	22	43	56	72	84	94	100	91	77	50
30	0	-5	-8	-10	-9	8	28	39	51	67	81	91	100	90	71
20	-3	-10	-14	-20	-21	-7	12	23	30	45	63	77	90	100	87
10	-11	-20	-25	-30	-33	-23	3	10	10	17	36	50	71	87	100

Table 6. Vertical correlations of U component of the wind between mandatory levels.

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
							latitude < -60°								
1000	100	90	89	56	48	44	40	35	19	-47	-70	-72	-95	-	-
850	90	100	98	79	67	60	52	49	25	-19	-88	-88	-76	-	-
700	89	98	100	84	72	66	60	55	30	-4	-76	-75	-61	-	-
500	56	79	84	100	94	91	83	69	39	18	-12	-10	16	-	-
400	48	67	72	94	100	96	85	63	45	20	-19	-16	69	-	-
300	44	60	66	91	96	100	95	79	51	29	19	21	78	-	-
250	40	52	60	83	85	95	100	92	71	60	55	57	69	-	-
200	35	49	55	69	63	79	92	100	92	89	89	90	46	-	-
150	19	25	30	39	45	51	71	92	100	99	98	97	48	-	-
100	-47	-19	-4	18	20	29	60	89	99	100	99	98	42	-	-
70	-70	-88	-76	-12	-19	19	55	89	98	99	100	99	90	-	-
50	-72	-88	-75	-10	-16	21	57	90	97	98	99	100	95	85	80
30	-95	-76	-61	16	69	78	69	46	48	42	90	95	100	90	85
20	-	-	-	-	-	-	-	-	-	-	-	85	90	100	90
10	-	-	-	-	-	-	-	-	-	-	-	80	85	90	100
							-60° < latitude < -30°								
1000	100	92	88	72	66	63	61	58	53	41	9	-12	4	32	-
850	92	100	95	82	76	74	73	71	67	53	18	-11	-1	31	-
700	88	95	100	94	90	89	87	82	75	59	21	-13	-17	15	-
500	72	82	94	100	98	97	93	84	73	55	21	-13	-22	6	-
400	66	76	90	98	100	98	95	86	76	57	23	-14	-23	6	-
300	63	74	89	97	98	100	98	92	82	64	30	-9	-26	6	-
250	61	73	87	93	95	98	100	97	90	74	41	1	-19	22	-
200	58	71	82	84	86	92	97	100	97	85	51	18	-20	49	-
150	53	67	75	73	76	82	90	97	100	94	64	38	21	44	-
100	41	53	59	55	57	64	74	85	94	100	85	72	67	47	-
70	9	18	21	21	23	30	41	51	64	85	100	90	79	70	-
50	-12	-11	-13	-13	-14	-9	1	18	38	72	90	100	91	83	-
30	4	-1	-17	-22	-23	-26	-19	-20	21	67	79	91	100	97	85
20	32	31	15	6	6	6	22	49	44	47	70	83	97	100	90
10	-	-	-	-	-	-	-	-	-	-	-	-	85	90	100

Table 6. Continued

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
							-30° < latitude < 30°								
1000	100	80	73	31	7	-5	-8	-5	-2	1	-7	-13	-25	-35	-66
850	80	100	86	49	28	17	14	16	17	20	5	-2	-9	-20	-61
700	73	86	100	74	58	49	44	41	41	42	33	28	20	10	-36
500	31	49	74	100	93	81	75	69	66	60	53	40	33	24	2
400	7	28	58	93	100	93	88	82	77	69	63	53	52	43	22
300	-5	17	49	81	93	100	97	92	87	77	71	62	60	51	27
250	-8	14	44	75	88	97	100	97	94	83	74	66	62	52	49
200	-5	16	41	69	82	92	97	100	97	88	77	70	64	53	64
150	-2	17	41	66	77	87	94	97	100	93	84	77	70	60	75
100	1	20	42	60	69	77	83	88	93	100	95	87	79	68	69
70	-7	5	33	53	63	71	74	77	84	95	100	96	91	85	80
50	-13	-2	28	40	53	62	66	70	77	87	96	100	97	93	91
30	-25	-9	20	33	52	60	62	64	70	79	91	97	100	98	95
20	-35	-20	10	24	43	51	52	53	60	68	85	98	98	100	96
10	-66	-61	-36	2	22	27	49	64	75	69	80	91	95	96	100
							30° < latitude < 60°								
1000	100	90	84	72	63	50	43	35	32	24	21	21	15	14	5
850	90	100	93	79	69	55	47	38	34	28	25	25	18	18	12
700	84	93	100	92	85	73	67	59	51	43	38	36	28	29	25
500	72	79	92	100	97	90	85	78	69	59	50	46	38	40	38
400	63	69	85	97	100	96	93	86	77	65	55	51	43	45	43
300	50	55	73	90	96	100	98	93	83	71	59	53	47	47	44
250	43	47	67	85	93	98	100	97	88	75	62	56	49	50	47
200	35	38	59	78	86	93	97	100	95	83	71	64	56	59	55
150	32	34	51	69	77	83	88	95	100	93	84	77	71	71	68
100	24	28	43	59	65	71	75	83	93	100	96	91	87	84	77
70	21	25	38	50	55	59	62	71	84	96	100	97	94	91	84
50	21	25	36	46	51	53	56	64	77	91	97	100	98	95	90
30	15	18	28	38	43	47	49	56	71	87	94	98	100	98	93
20	14	18	29	40	45	47	50	59	71	84	91	95	98	100	96
10	5	12	25	38	43	44	47	55	68	77	84	90	93	96	100

Table 6. Continued

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
					latitude > 60°										
1000	100	85	79	62	53	54	52	48	45	34	28	13	7	24	11
850	85	100	95	83	75	67	65	66	62	44	34	19	12	27	10
700	79	95	100	95	91	84	82	82	73	51	36	18	9	21	-3
500	62	83	95	100	98	95	92	90	79	56	38	22	13	19	-3
400	53	75	91	98	100	98	96	93	80	57	38	23	12	16	-5
300	54	67	84	95	98	100	99	95	83	60	40	26	14	15	-7
250	52	65	82	92	96	99	100	97	85	64	45	31	19	17	-1
200	48	66	82	90	93	95	97	100	94	76	57	43	30	30	9
150	45	62	73	79	80	83	85	94	100	91	77	62	49	47	26
100	34	44	51	56	57	60	64	76	91	100	94	89	76	73	57
70	28	34	36	38	38	40	45	57	77	94	100	98	88	86	77
50	13	19	18	22	23	26	31	43	62	89	98	100	97	93	81
30	7	12	9	13	12	14	19	30	49	76	88	97	100	97	86
20	24	27	21	19	16	15	17	30	47	73	86	93	97	100	91
10	11	10	-3	-3	-5	-7	-1	9	26	57	77	81	86	91	100

Table 7. Vertical correlations of V component of the wind between mandatory levels.

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
							latitude < -60°								
1000	100	90	89	56	48	44	40	35	19	-47	-70	-72	-95	-	-
850	90	100	98	79	67	60	52	49	25	-19	-88	-88	-76	-	-
700	89	98	100	84	72	66	60	55	30	-4	-76	-75	-61	-	-
500	56	79	84	100	94	91	83	69	39	18	-12	-10	16	-	-
400	48	67	72	94	100	96	85	63	45	20	-19	-16	69	-	-
300	44	60	66	91	96	100	95	79	51	29	19	21	78	-	-
250	40	52	60	83	85	95	100	92	71	60	55	57	69	-	-
200	35	49	55	69	63	79	92	100	92	89	89	90	46	-	-
150	19	25	30	39	45	51	71	92	100	99	98	97	48	-	-
100	-47	-19	-4	18	20	29	60	89	99	100	99	98	42	-	-
70	-70	-88	-76	-12	-19	19	55	89	98	99	100	99	90	-	-
50	-72	-88	-75	-10	-16	21	57	90	97	98	99	100	95	85	80
30	-95	-76	-61	16	69	78	69	46	48	42	90	95	100	90	85
20	-	-	-	-	-	-	-	-	-	-	-	85	90	100	90
10	-	-	-	-	-	-	-	-	-	-	-	80	85	90	100
							-60° < latitude < -30°								
1000	100	92	88	72	66	63	61	58	53	41	9	-12	4	32	-
850	92	100	95	82	76	74	73	71	67	53	18	-11	-1	31	-
700	88	95	100	94	90	89	87	82	75	59	21	-13	-17	15	-
500	72	82	94	100	98	97	93	84	73	55	21	-13	-22	6	-
400	66	76	90	98	100	98	95	86	76	57	23	-14	-23	6	-
300	63	74	89	97	98	100	98	92	82	64	30	-9	-26	6	-
250	61	73	87	93	95	98	100	97	90	74	41	1	-19	22	-
200	58	71	82	84	86	92	97	100	97	85	51	18	-20	49	-
150	53	67	75	73	76	82	90	97	100	94	64	38	21	44	-
100	41	53	59	55	57	64	74	85	94	100	85	72	67	47	-
70	9	18	21	21	23	30	41	51	64	85	100	90	79	70	-
50	-12	-11	-13	-13	-14	-9	1	18	38	72	90	100	91	83	-
30	4	-1	-17	-22	-23	-26	-19	-20	21	67	79	91	100	97	85
20	32	31	15	6	6	6	22	49	44	47	70	83	97	100	90
10	-	-	-	-	-	-	-	-	-	-	-	-	85	90	100

Table 7. Continued

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
-30° < latitude < 30°															
1000	100	80	73	31	7	-5	-8	-5	-2	2	-7	-13	-25	-35	-66
850	80	100	86	49	28	17	14	16	17	20	5	-2	-9	-20	-61
700	73	86	100	74	58	49	44	41	41	42	33	28	20	10	-36
500	31	49	74	100	93	81	75	69	66	60	53	40	33	24	2
400	7	28	58	93	100	93	88	82	77	69	63	53	52	43	22
300	-5	17	49	81	93	100	97	92	87	77	71	62	60	51	27
250	-8	14	44	75	88	97	100	97	94	83	74	66	62	52	49
200	-5	16	41	69	82	92	97	100	97	88	77	70	64	53	64
150	-2	17	41	66	77	87	94	97	100	93	84	77	70	60	75
100	2	20	42	60	69	77	83	88	93	100	95	87	79	68	69
70	-7	5	33	53	63	71	74	77	84	95	100	96	91	85	80
50	-13	-2	28	40	53	62	66	70	77	87	96	100	97	93	91
30	-25	-9	20	33	52	60	62	64	70	79	91	97	100	98	95
20	-35	-20	10	24	43	51	52	53	60	68	85	93	98	100	96
10	-66	-61	-36	2	22	27	49	64	75	69	80	91	95	96	100
30° < latitude < 60°															
1000	100	90	84	72	63	50	43	35	32	24	21	21	14	14	5
850	90	100	93	79	69	55	47	38	34	28	25	25	18	18	12
700	84	93	100	92	85	73	67	59	51	43	38	36	28	29	25
500	72	79	92	100	97	90	85	78	69	59	50	46	38	40	38
400	63	69	85	97	100	96	93	86	77	65	55	51	43	45	43
300	50	55	73	90	96	100	98	93	83	71	59	53	47	47	44
250	43	47	67	85	93	98	100	97	88	75	62	56	49	50	47
200	35	38	59	78	86	93	97	100	95	83	71	64	56	59	55
150	32	34	51	69	77	83	88	95	100	93	84	77	71	71	68
100	24	28	43	59	65	71	75	83	93	100	96	91	87	84	77
70	21	25	38	50	55	59	62	71	84	96	100	97	94	91	84
50	21	25	36	46	51	53	56	64	77	91	97	100	98	95	90
30	14	18	28	38	43	47	49	56	71	87	94	98	100	98	93
20	14	18	29	40	45	47	50	59	71	84	91	95	98	100	96
10	5	12	25	38	43	44	47	55	68	77	84	90	93	96	100

Table 7. Continued

	1000	850	700	500	400	300	250	200	150	100	70	50	30	20	10
							latitude > 60°								
1000	100	85	79	62	53	44	42	41	40	34	28	13	7	24	11
850	85	100	95	83	75	67	65	66	62	44	34	19	12	27	10
700	79	95	100	95	91	84	82	82	73	51	36	18	9	21	-3
500	62	83	95	100	98	95	92	90	79	56	38	22	13	19	-3
400	53	75	91	98	100	98	96	93	80	57	38	23	12	16	-5
300	44	67	84	95	98	100	99	95	83	60	40	26	14	15	-7
250	42	65	82	92	96	99	100	97	85	64	45	31	19	17	-1
200	41	66	82	90	93	95	97	100	94	76	57	43	30	30	9
150	40	62	73	79	80	83	85	94	100	91	77	62	49	47	26
100	34	44	51	56	57	60	64	76	91	100	94	89	76	73	57
70	28	34	36	38	38	40	45	57	77	94	100	98	88	86	77
50	13	19	18	22	23	26	31	43	62	89	98	100	97	93	81
30	7	12	9	13	12	14	19	30	49	76	88	97	100	97	86
20	24	27	21	19	16	15	17	30	47	73	86	93	97	100	91
10	11	10	-3	-3	-5	-7	-1	9	26	57	77	81	86	91	100

Table 8. Vertical correlations of dewpoint depression between mandatory levels.

latitude < -60°	1000	850	700	500	400	300
1000	100	40	30	20	18	15
850	40	100	39	34	30	6
700	30	39	100	64	55	7
500	20	34	64	100	79	8
400	18	30	55	79	100	50
300	15	6	7	8	50	100
-60° < latitude < -30°						
1000	100	-2	49	40	38	48
850	-2	100	0	-34	-19	-42
700	49	0	100	51	32	31
500	40	-34	51	100	72	70
400	38	-19	32	72	100	79
300	48	-42	31	70	79	100
-30° < latitude < 30°						
1000	100	5	0	-8	-22	-25
850	5	100	32	23	27	39
700	0	32	100	30	45	39
500	-8	23	30	100	59	38
400	-22	27	45	59	100	71
300	-25	39	39	38	71	100
30° < latitude < 60°						
1000	100	39	32	29	37	29
850	39	100	45	30	26	16
700	32	45	100	47	41	42
500	29	30	47	100	83	69
400	37	26	41	83	100	85
300	29	16	42	69	85	100
latitude > 60°						
1000	100	52	23	32	27	23
850	52	100	32	22	1	22
700	23	32	100	38	54	58
500	32	22	38	100	73	83
400	27	1	54	73	100	88
300	23	22	58	83	88	100

Table 9. CQC component response to different types of gross errors in geopotential height and temperature at mandatory levels. The first character is: H represents horizontal, V represents vertical, S represents the use of significant levels; the second character is: H represent geopotential height, h thickness, T temperature, and "**" represents the hydrostatic check. In the table '+' indicates the presence of response, '±' the presence of weak response, '-' the absent of a response, B_i^j the corresponding hydrostatic check coefficients; and a_i^j the corresponding coefficients of the vertical check.

Type of error	Level	**	HH	Hh	VH	HT	VT	ST
	.							
Error χ in H_1 at lowest level	3		-		-	-	-	-
		-		-				
	2		-		$-a_1^2\chi$	-	-	-
		$-\chi$		$-\chi$				
	1		χ		χ	-	-	-
	.							
Error χ in H_i at intermediate level	i+1		-		$-a_i^{i+1}\chi$	-	-	-
		$-\chi$		$-\chi$				
	i		χ		χ	-	-	-
		χ		χ				
	i-1		-		$-a_i^{i-1}\chi$	-	-	-
	.							
Error χ in H_n at upper level	n		χ		χ	-	-	-
		χ		χ				
	n-1		-		$-a_n^{n-1}\chi$	-	-	-
		-		-				
	n-2		-		-	-	-	-
	.							
	.							
Error τ in T_1 at lowest level	3		-		-	-	-	-
		-		-				
	2		-		-	-	$-a_1^2\tau$	-
		$-B_1^2\tau$		-				
	1		-		-	τ	τ	τ
	.							
Error τ in T_i at intermediate level	i+1		-		-	-	$-a_i^{i+1}\tau$	-
		$-B_i^{i+1}\tau$		-				
	i		-		-	τ	τ	τ
		$-B_i^{i-1}\tau$		-				
	i-1		-		-	-	$-a_i^{i-1}\tau$	-
	.							

Table 9. Continued

Type of error	Level	**	HH	Hh	VH	HT	VT	ST
Error τ in T_n at upper level	n		-		-	τ	τ	τ
		$-B_n^{n-1}\tau$		-				
	n-1		-		-	-	$-a_n^{n-1}\tau$	-
		-		-				

	n-2		-		-	-	-	-
	.							
	.							
Miscalculation χ starting at H_1	3		χ		-	-	-	-
		-		-				
	2		χ		±	-	-	-
		-		-				
	1		χ		±	-	-	-
	.							
Miscalculation χ starting at H_i	i+1		χ		-	-	-	-
		-		-				
	i		χ		±	-	-	-
		χ		χ				
	i-1		-		$-a_i^{i+1}\chi$	-	-	-
	.							
Radiosonde malfunction stating at i-th level	i-2		+		±	+	±	-
		-		±				
	i+1		+		±	±	±	-
		-		±				
	i		±		±	±	±	-
		-		±				
	i-1		-		±	-	-	-
	.							
	.							
		-						
	4							
		-						
Wrong station coordinates	3		±		±	±	±	-
		-		±				
	2		+		±	+	±	-
		-		+				
	1		+		±	+	±	-

Table 10. Distribution of CQC quality flags for geopotential heights (H) and temperatures (T) at mandatory levels for a global set of upper-air data for 1985/01/15/00.

	INPUT DATA		MANDATORY LEVELS, H & T, 1985/01/15/00, 748 STATIONS											
			UNCHECKED		CORRECT		SUSPECTED		ERRONEOUS		CORRCTED		CALCULATED	
1000	422	377	0	0	415	375	6	1	0	0	1	1	0	0
850	659	651	0	0	648	646	4	2	1	1	6	4	0	6
700	673	674	0	0	662	667	5	4	2	2	4	1	1	1
500	680	679	0	0	666	674	6	2	3	3	5	1	1	1
400	645	648	0	0	636	644	2	1	3	3	4	0	4	6
300	637	638	0	0	626	631	1	2	4	4	6	2	6	7
250	624	630	0	0	615	624	0	0	4	5	5	1	6	6
200	618	622	0	0	605	617	1	0	5	6	7	0	0	1
150	596	603	0	0	586	596	1	0	5	6	4	1	2	3
100	578	587	0	0	568	580	3	2	4	5	4	0	4	0
70	497	496	0	0	484	489	4	0	4	4	5	3	1	2
50	464	465	0	0	447	459	7	0	5	5	5	1	0	1
30	391	393	0	0	384	389	0	1	3	3	4	0	0	1
20	311	310	0	0	302	305	1	0	3	3	5	2	0	2
10	106	106	0	0	103	104	1	1	0	0	2	1	0	0
Tot.	7901	7879	0	0	7747	7800	42	16	46	50	67	18	25	37

Table 11. Distribution of CQC quality flags for wind speed and direction at mandatory levels for a global set of upper-air data for 1985/01/15/00.

P	INPUT		MAND. LEVELS, WIND, 1985/01/15/00, 748 STAT.							
	DATA		UNCHECKED		CORRECT		SUSPECTED		ERRONEOUS	
1000	269	269	0	0	263	263	0	0	6	6
850	665	665	0	0	632	640	5	5	27	19
700	701	701	0	0	668	672	6	6	25	21
500	706	706	0	0	683	687	4	4	16	12
400	654	654	0	0	641	641	3	3	7	7
300	630	630	0	0	619	620	2	2	5	4
250	610	610	0	0	599	601	2	2	5	3
200	595	595	0	0	584	583	1	1	6	7
150	565	565	0	0	551	551	2	2	8	8
100	530	530	0	0	512	516	2	2	11	7
70	451	451	0	0	436	438	1	1	10	8
50	419	419	0	0	396	400	5	5	13	9
30	338	338	0	0	326	323	0	0	9	12
20	256	256	0	0	250	252	0	0	4	2
10	72	72	0	0	70	70	0	0	2	2
Tot.	7461	7461	0	0	7230	7257	33	33	154	127

Table 12. Distribution of CQC quality flags for dewpoint depression at mandatory levels for global upper-air observations 1985/01/15/00.

P	INPUT	MAND. LEV., HUMID., 1985/01/15/00, 748 STAT.			
	DATA	UNCHECKED	CORRECT	SUSPECTED	CORRECTED
1000	321	0	320	0	1
850	536	1	531	1	2
700	547	5	532	1	7
500	528	1	520	1	4
400	451	0	446	1	1
300	398	1	395	0	1
250	315	315	0	0	0
200	296	296	0	0	0
150	244	244	0	0	0
100	201	201	0	0	0
70	136	136	0	0	0
50	125	125	0	0	0
30	99	99	0	0	0
20	77	77	0	0	0
10	21	21	0	0	0
Tot.	4295	1522	2744	4	16

Table 13. Distribution of CQC quality flags for geopotential heights and temperatures at significant levels for global upper-air observations 1985/01/15/00.

P1 - P2	INPUT		SIGNIF. LEVELS, H & T, 1985/01/15/00, 748 STAT.							
	DATA		UNCHECKED		CORRECT		SUSPECTED		ERRONEOUS	
< 1000	116	426	0	3	115	422	0	0	1	1
1000-850	689	1661	10	12	678	1635	0	2	1	12
850-700	621	1712	0	16	621	1691	0	0	0	5
700-500	890	1966	0	23	890	1933	0	0	0	10
500-400	317	788	0	7	317	774	0	0	0	7
400-300	278	840	0	12	278	821	0	0	0	7
300-250	99	479	0	4	99	472	0	0	0	3
250-200	185	594	0	4	185	581	0	0	0	9
200-150	355	726	0	5	355	718	0	0	0	3
150-100	329	706	0	1	329	695	0	0	0	10
100- 70	223	678	0	8	223	663	0	1	0	6
70- 50	197	499	0	5	197	488	0	0	0	6
50- 30	245	633	0	5	245	622	0	1	0	5
30- 20	152	379	0	2	152	374	0	0	0	3
20- 10	186	450	1	2	183	439	1	1	1	8
> 10	44	125	44	0	0	123	0	0	0	2
Total :	4926	12662	55	109	4867	12451	1	5	3	97

Table 14. Distribution of CQC quality flags for wind speed and direction at significant levels for global upper-air observations 1985/01/15/00.

P1 - P2	INPUT DATA		SIGNIF. LEVELS, WIND, 1985/01/15/01, 748 STAT. UNCHECKED		CORRECT		SUSPECTED		ERRONEOUS	
< 1000	348	348	80	80	268	268	0	0	0	0
1000-850	1059	1059	34	34	1024	1024	0	0	1	1
850-700	950	950	17	17	933	933	0	0	0	0
700-500	1221	1221	25	25	1191	1191	0	0	5	5
500-400	498	498	12	12	486	486	0	0	0	0
400-300	590	590	21	21	567	567	0	0	2	2
300-250	367	367	13	13	348	348	2	2	4	4
250-200	447	447	10	10	429	429	3	3	5	5
200-150	555	555	10	10	537	537	0	0	8	8
150-100	534	534	9	9	516	516	6	6	3	3
100- 70	381	381	14	14	366	366	0	0	2	1
70- 50	269	269	2	2	266	266	0	0	1	1
50- 30	347	347	13	13	332	332	0	0	2	2
30- 20	232	232	5	5	224	224	1	1	2	2
20- 10	212	212	0	0	211	211	0	0	1	1
> 10	41	41	0	0	40	40	0	0	1	1
Total :	8051	8051	265	265	7739	7760	11	6	37	37

Table 15. Distribution of CQC quality flags for dewpoint depression at significant levels for global set of upper-air data for 85/01/15/00.

P1 - P2	INPUT	SIGN. LEVELS, HUMID.,1985/01/15/00,748 STAT.			
	DATA	UNCHECKED	CORRECT	SUSPECTED	ERRONEOUS
< 1000	360	6	354	0	0
1000-850	1150	21	1129	0	0
850-700	1089	24	1063	0	2
700-500	1049	26	1017	0	6
500-400	436	9	422	0	5
400-300	464	15	445	0	4
300-250	267	206	60	0	1
250-200	249	248	0	0	1
200-150	201	200	0	0	1
150-100	171	168	0	0	3
100- 70	170	167	0	0	3
70- 50	110	109	0	0	1
50- 30	111	111	0	0	0
30- 20	69	67	0	0	2
20- 10	84	81	0	0	3
> 10	26	26	0	0	0
Total :	6006	1484	4490	0	32

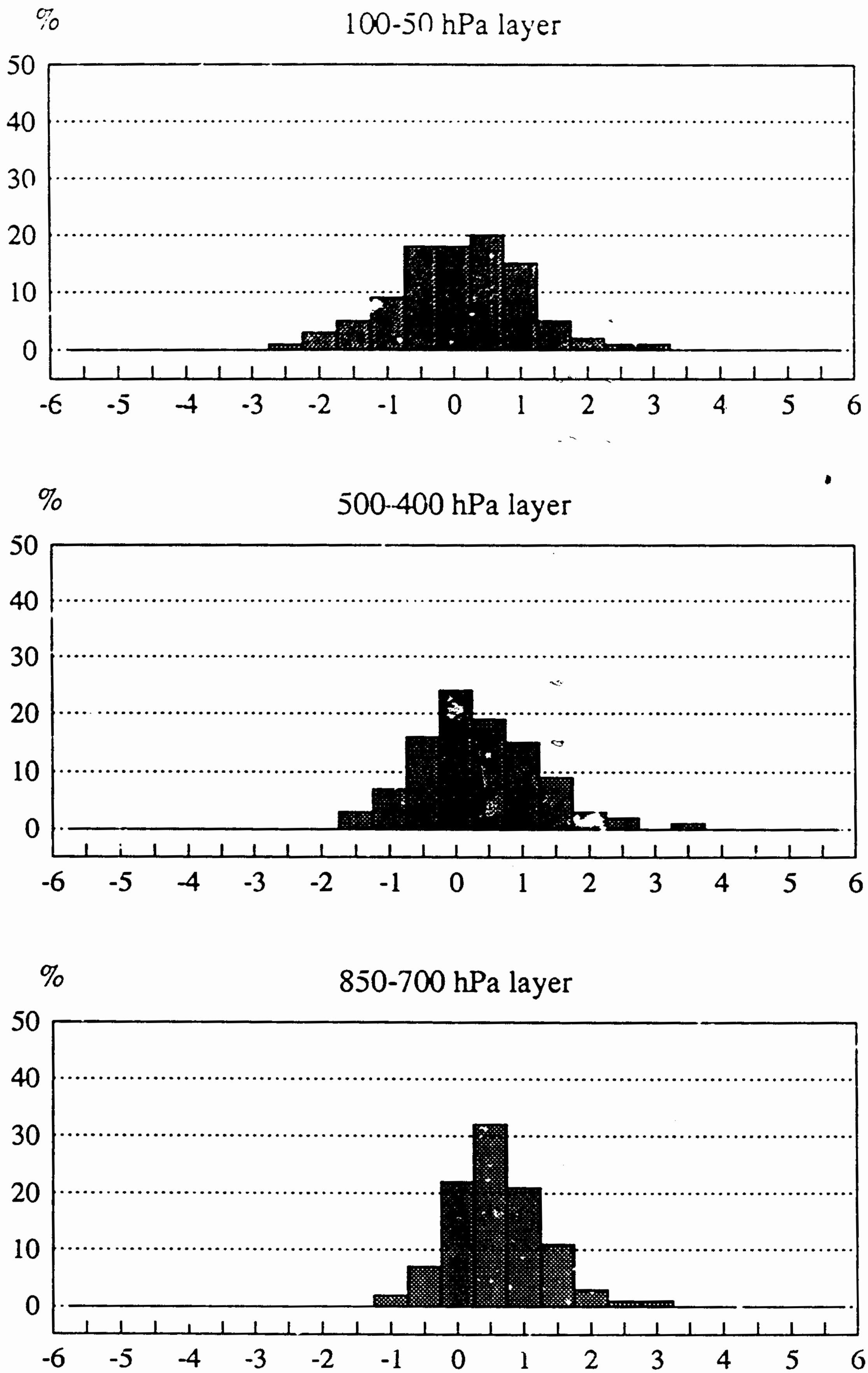


Fig. 1. Distribution of the normalized actual residuals from the hydrostatic check from a dataset of 759 stations from 00 UTC, 15 Jan 1989.

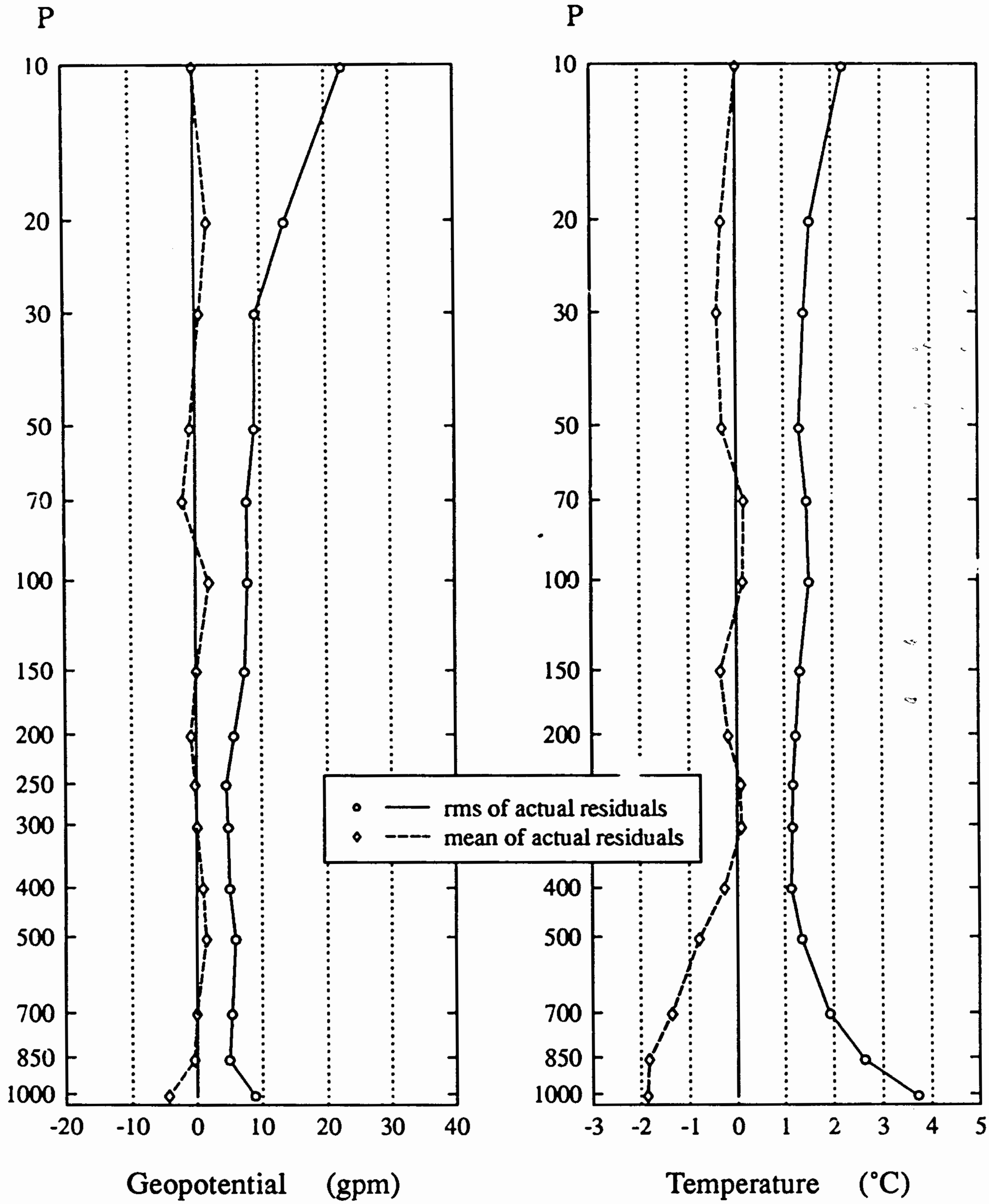


Fig. 2. Mean and RMS residuals for geopotential height and temperature calculated using the hydrostatic equation are shown for a global dataset from 00 UTC, 15 Jan 1989.

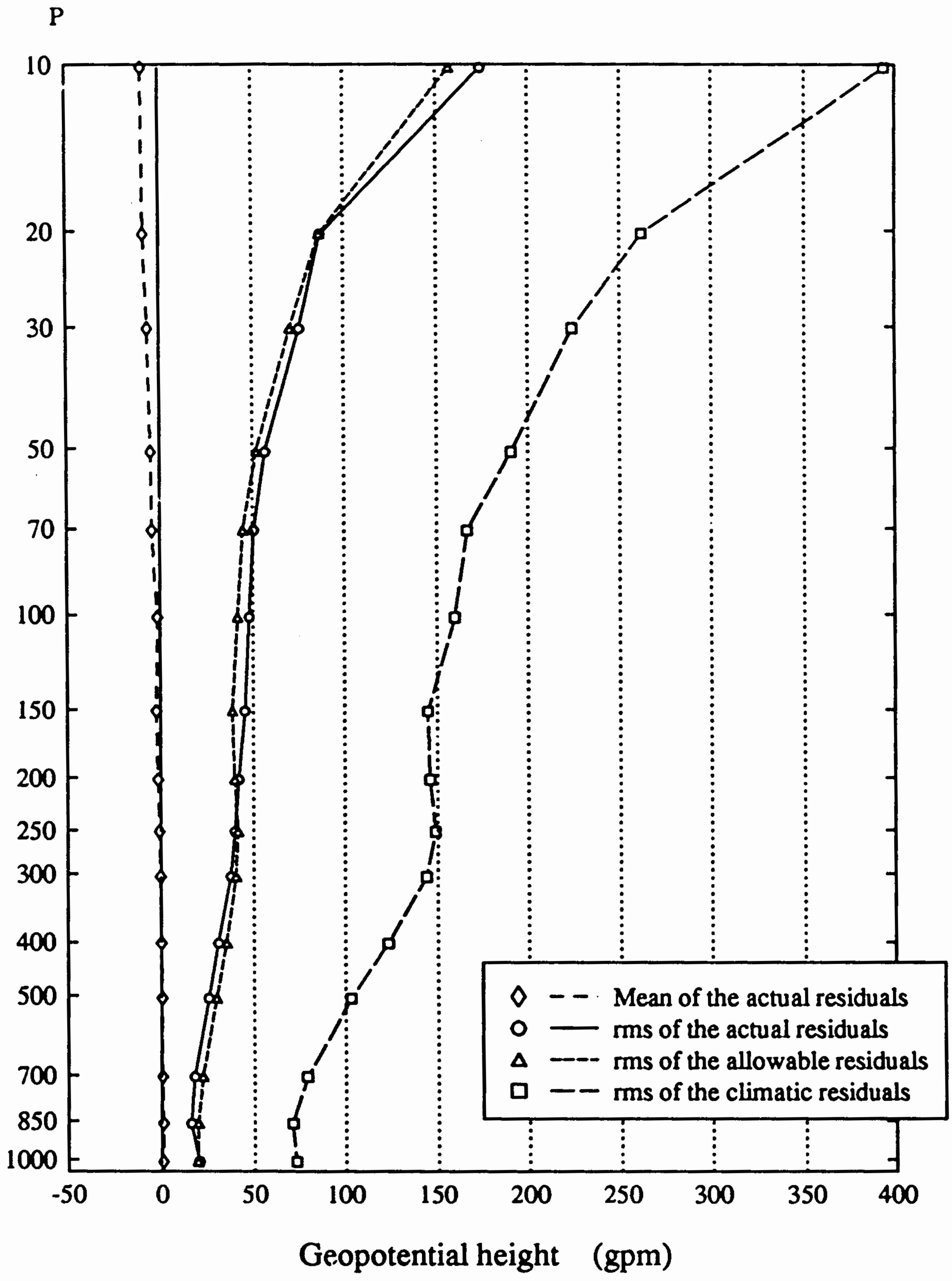


Fig. 3. Characteristics of the horizontal optimal interpolation of geopotential height from a global set of stations from 00 UTC 15 Jan 1989.

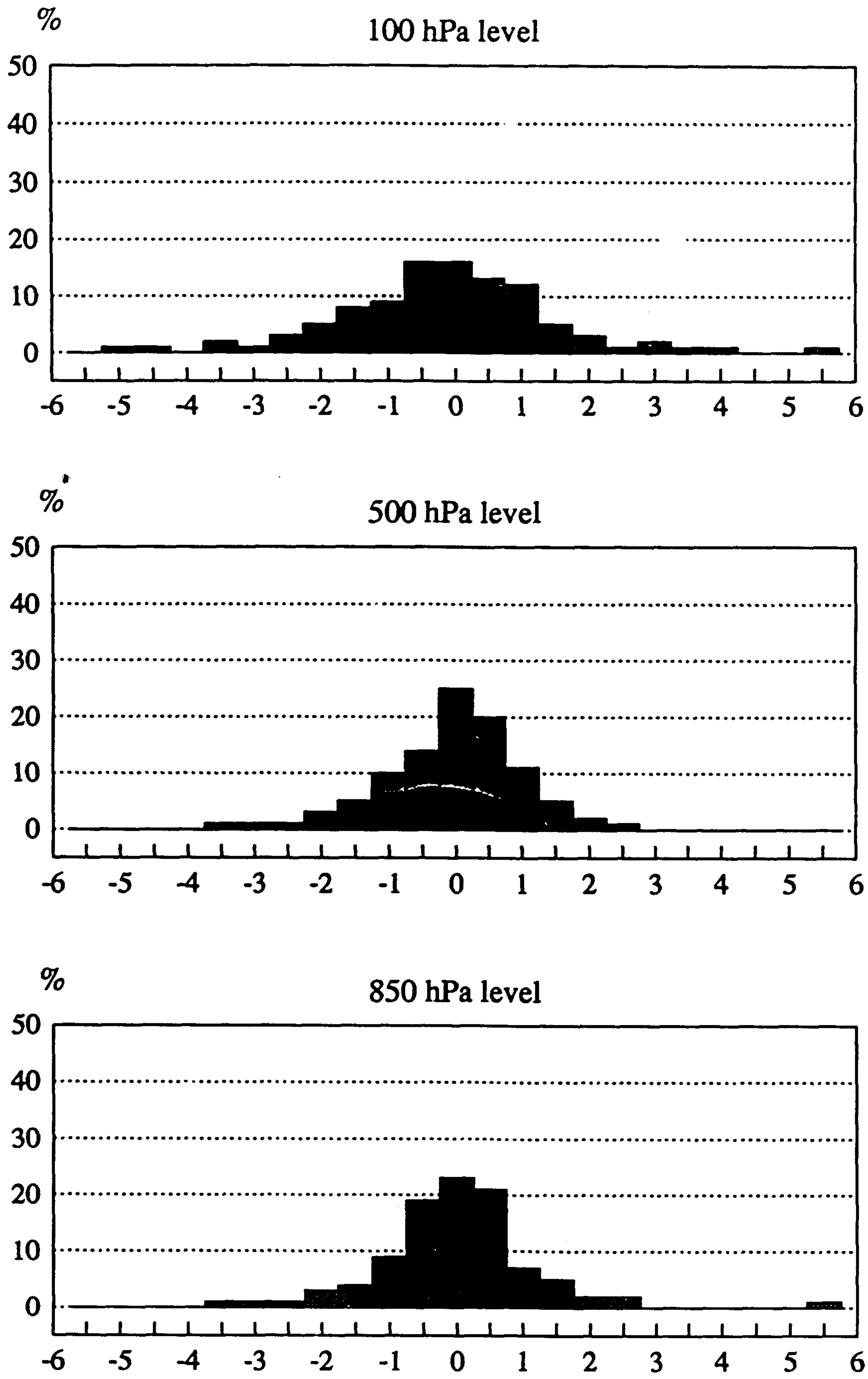


Fig. 4. Distribution of the normalized actual residuals for horizontal optimal interpolation of geopotential height for a global set of data from 00 UTC, 15 Jan 1989.

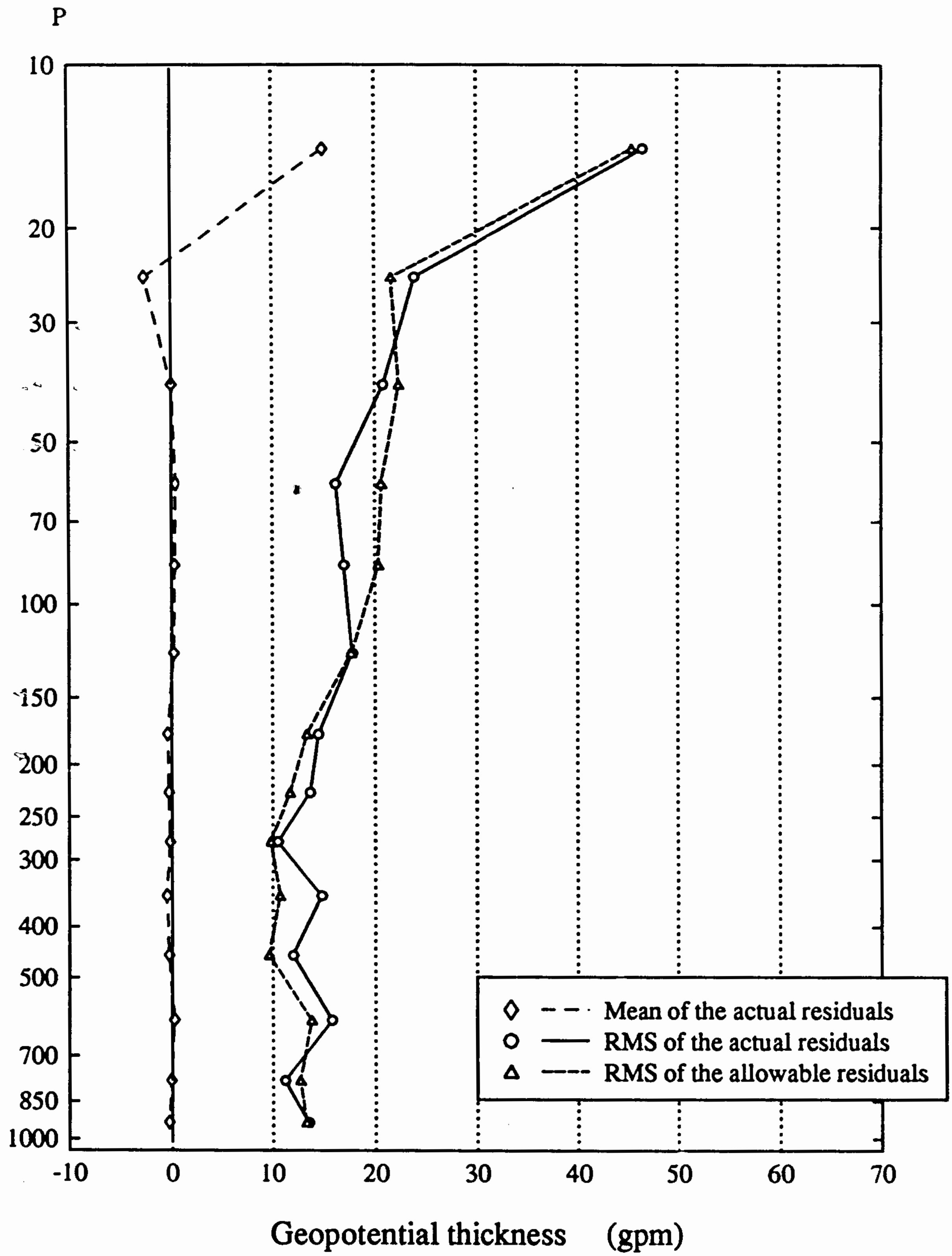


Fig. 5. Characteristics of horizontal optimal interpolation of geopotential thickness for a dataset of 759 station from 00 UTC, 15 Jan 1989.

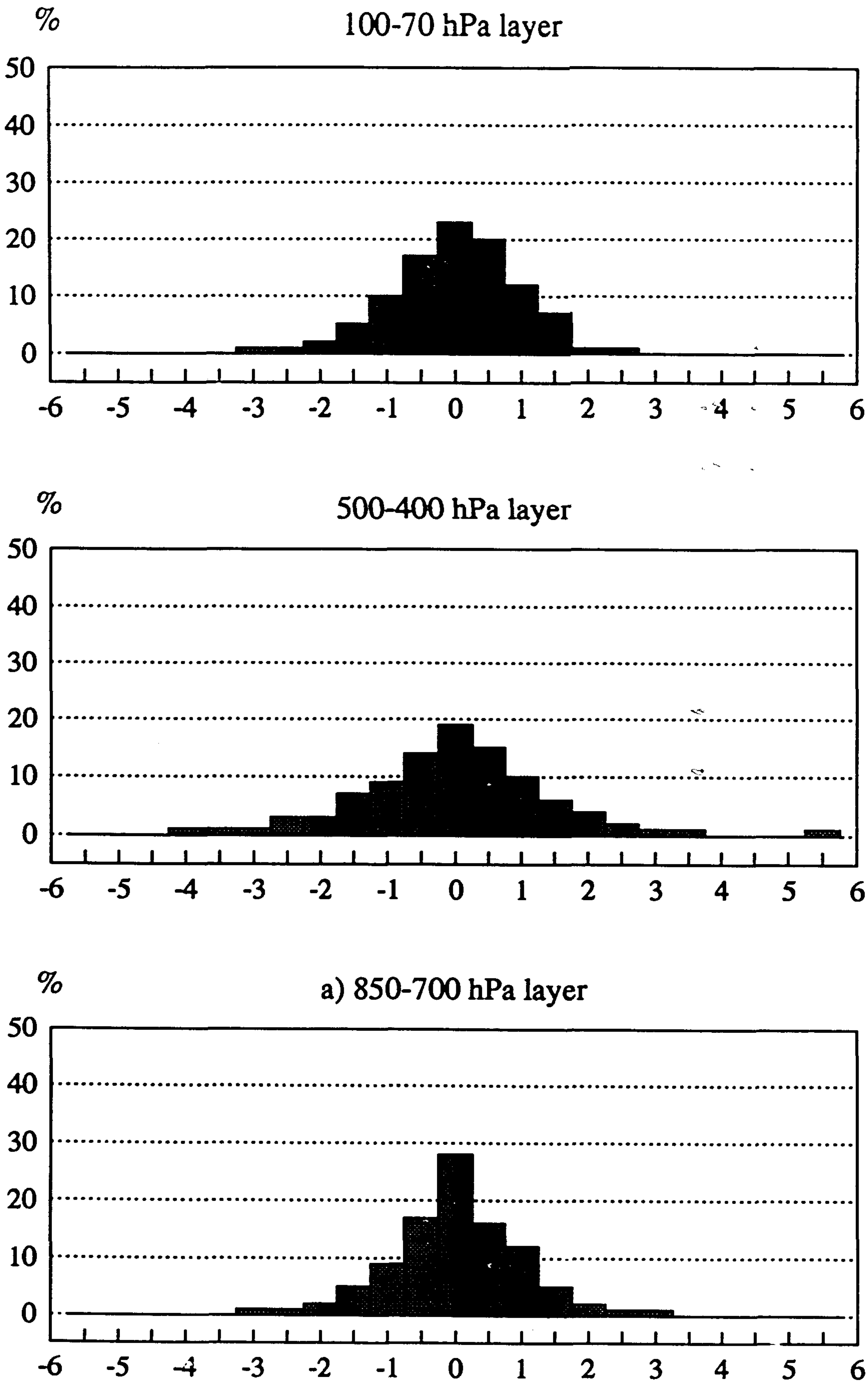


Fig. 6. Distribution of the normalized actual residuals from the horizontal optimal interpolation of geopotential thickness from a dataset of 759 stations from 00 UTC, 15 Jan 1989.

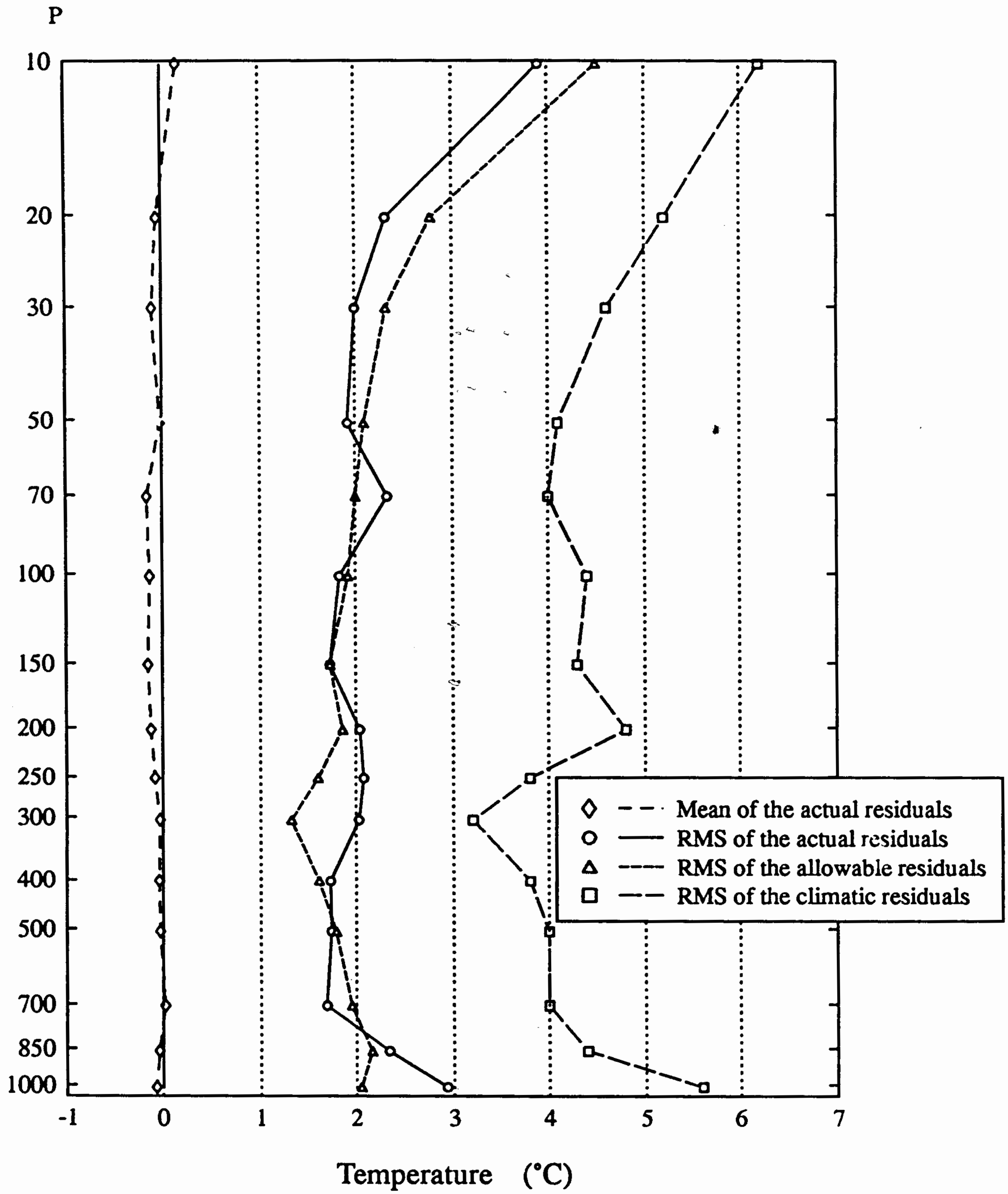


Fig. 7. Characteristics of the horizontal optimal interpolation of temperature for a dataset of 759 station from 00 UTC, 15 Jan 1989.

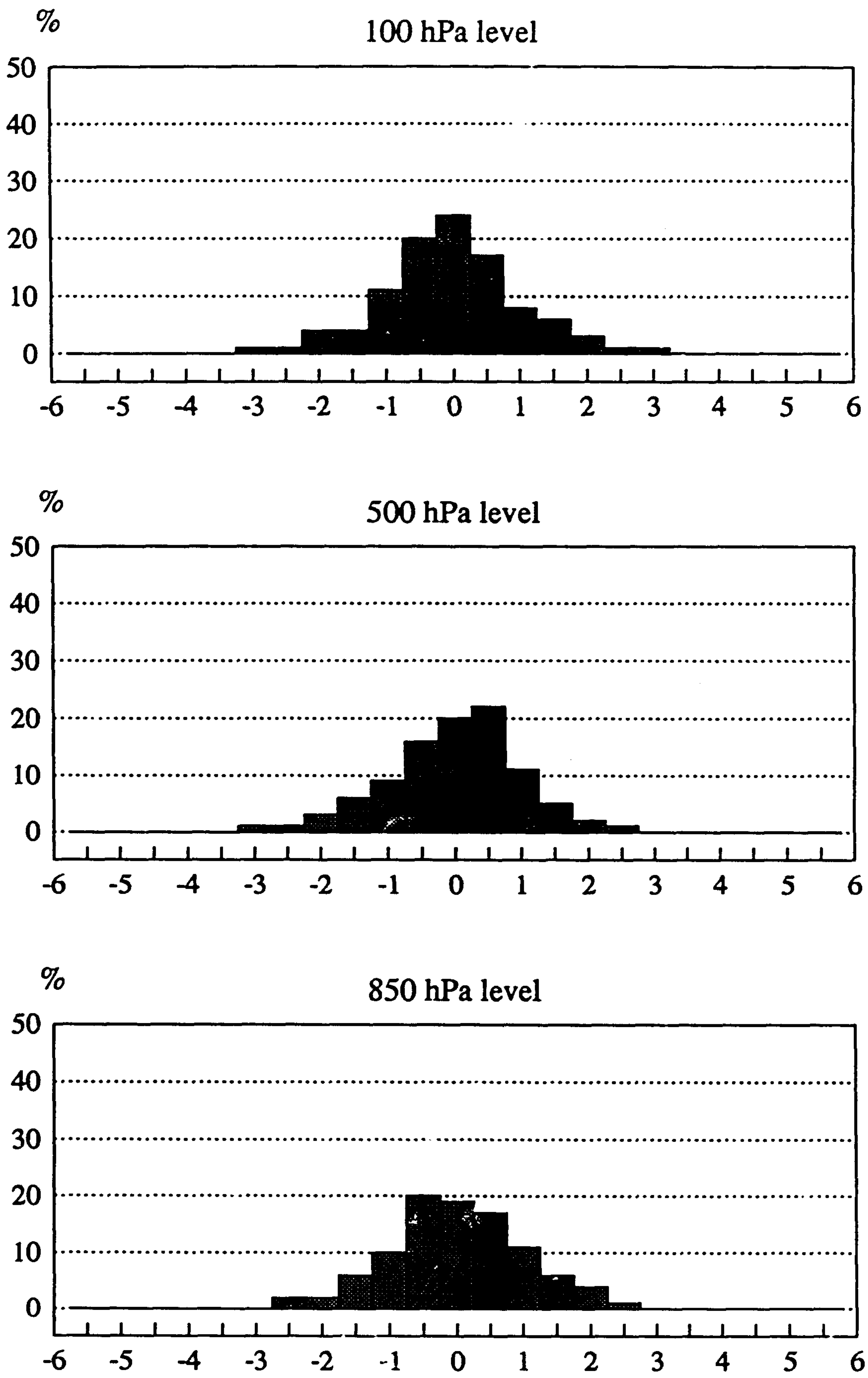


Fig. 8. Distribution of the normalized actual residuals from the horizontal optimal interpolation of temperature for a dataset of 759 stations from 00 UTC, 15 Jan 1989.

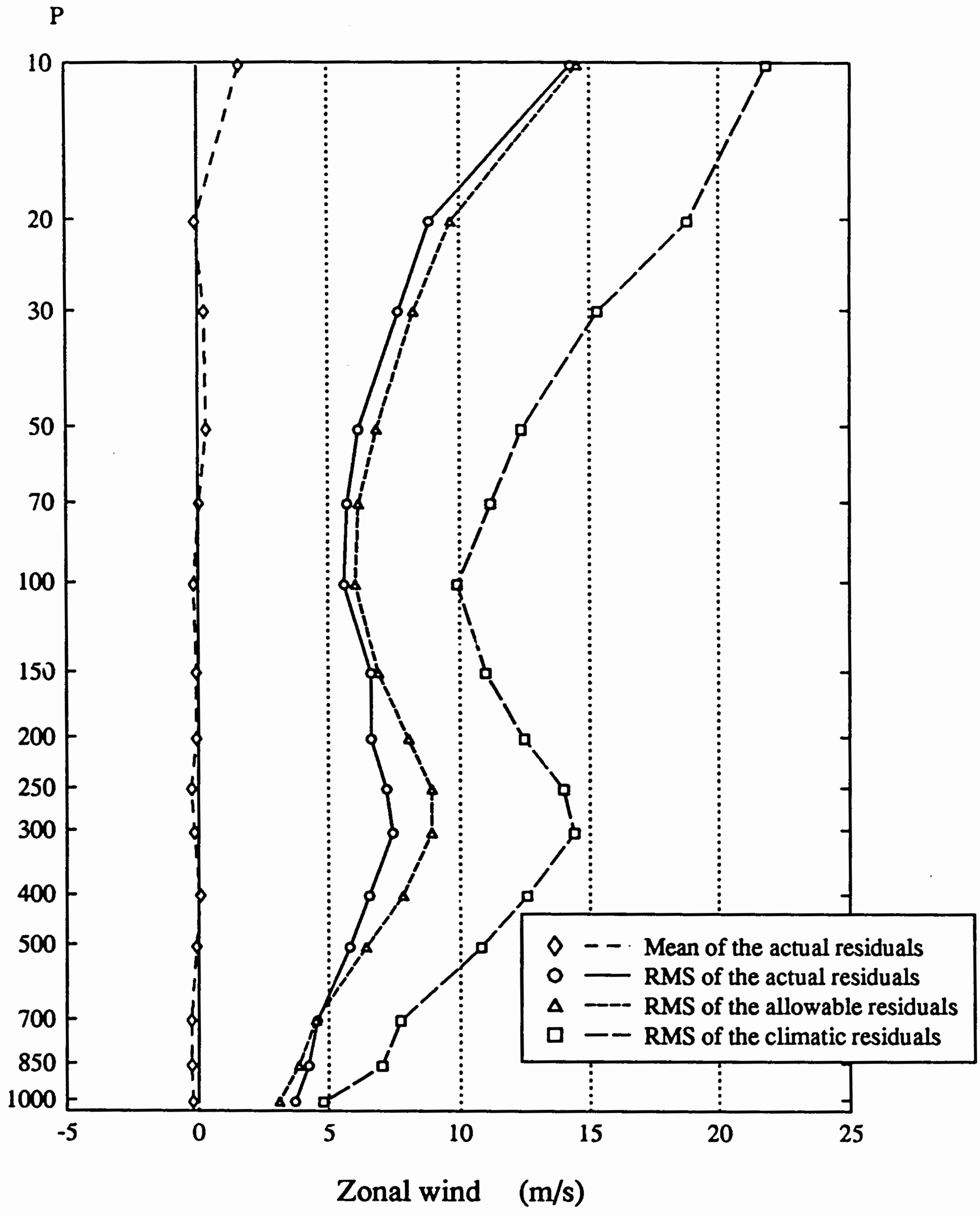


Fig. 9. Characteristics of the horizontal optimal interpolation of the zonal wind for a dataset of 759 stations from 00 UTC 15 Jan 1989.

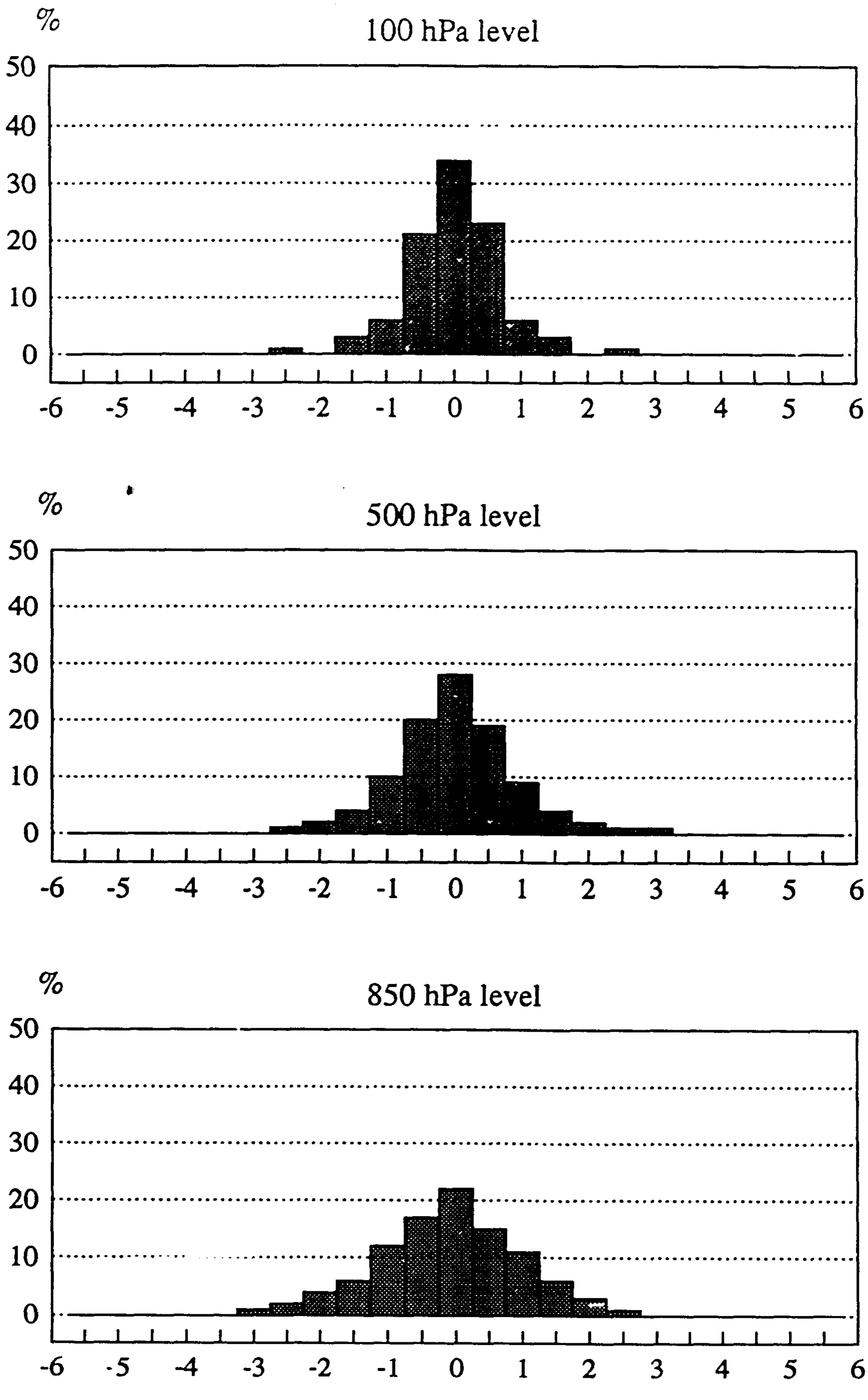


Fig. 10. Distribution of the normalized actual residuals from horizontal optimal interpolation of the U component of the wind for a dataset of 759 stations from 00 UTC 15 Jan 1989.

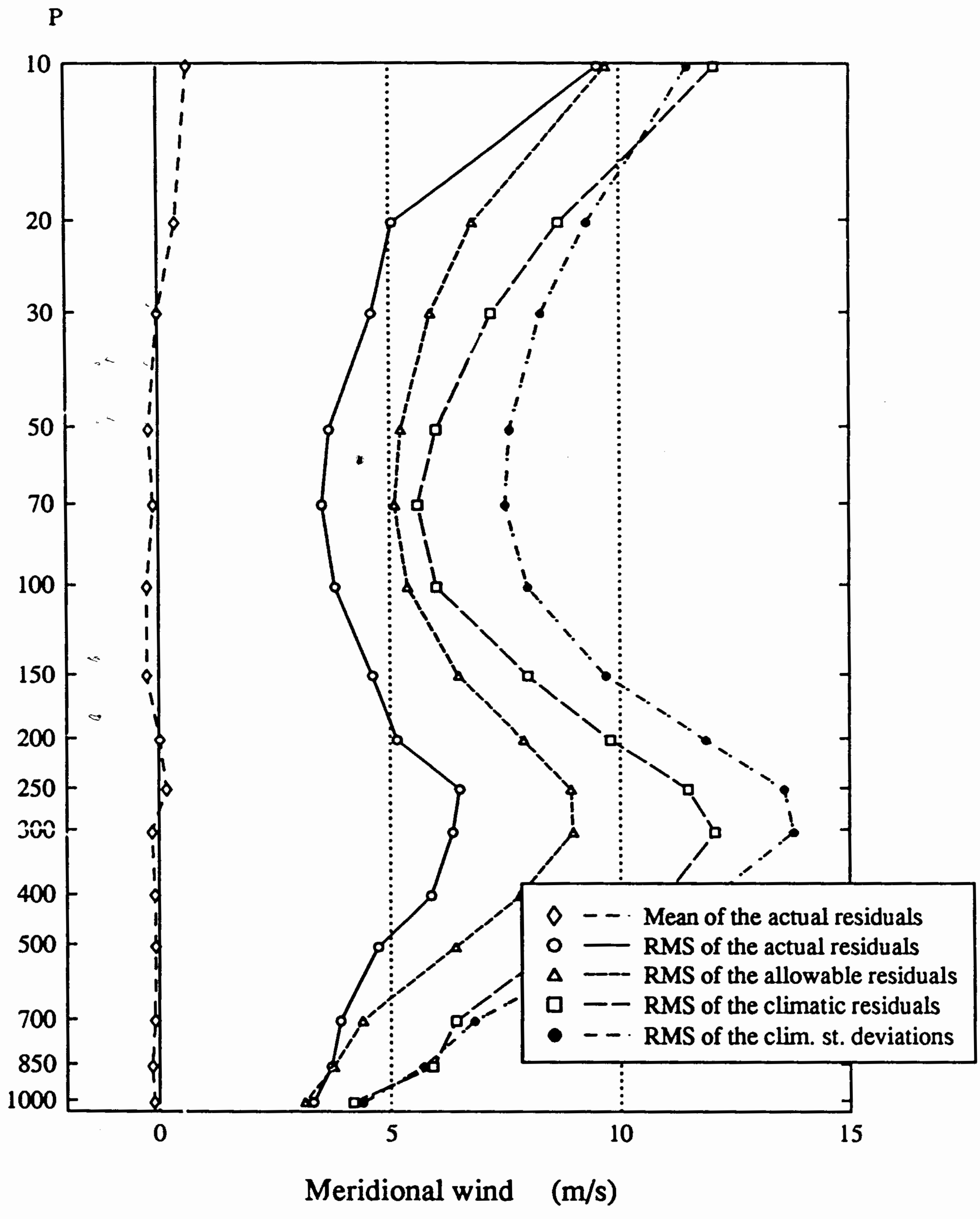
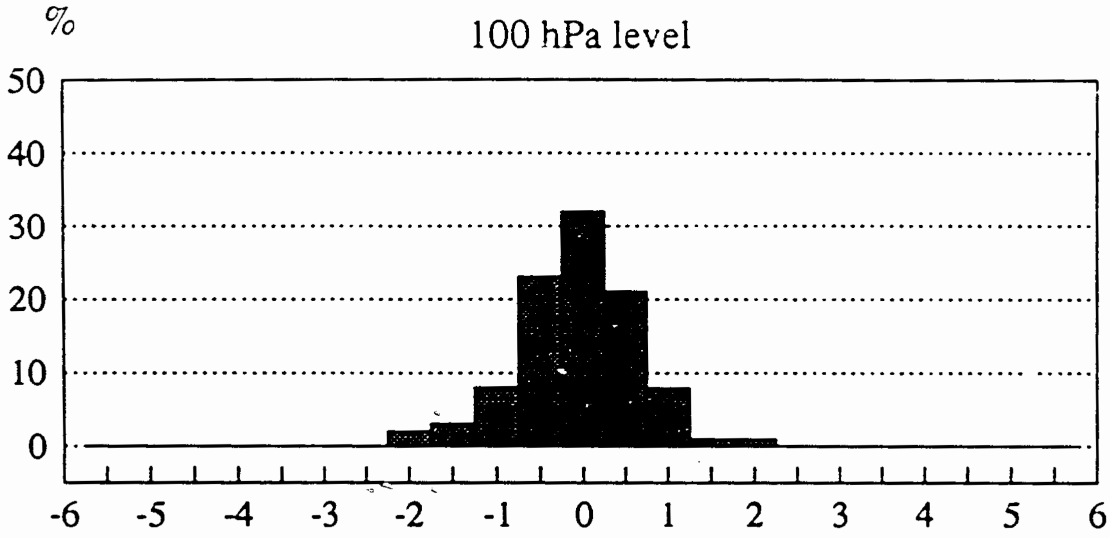
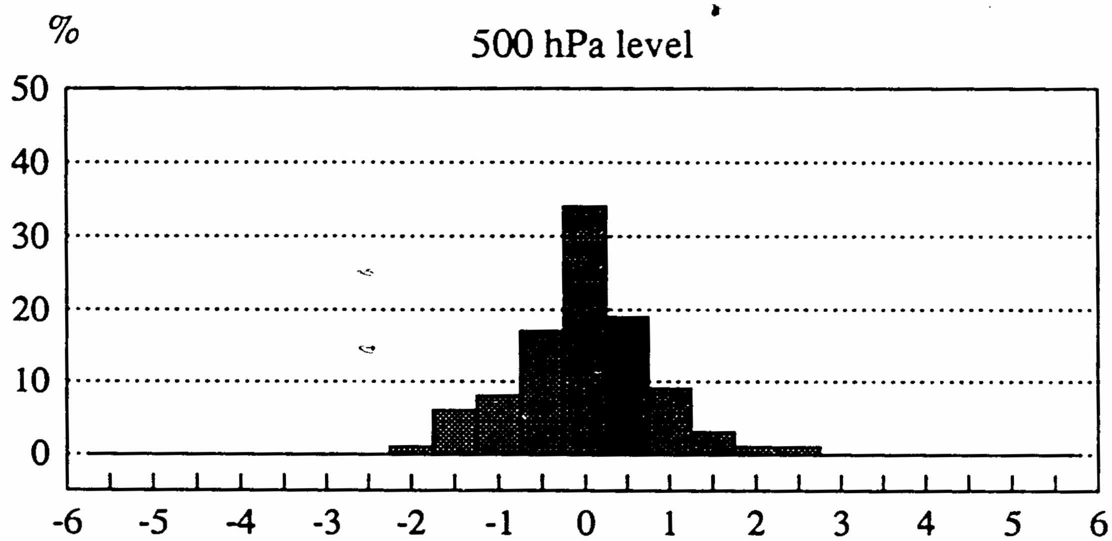


Fig. 11. Same as figure 3 except for the meridional (V) component of the wind.





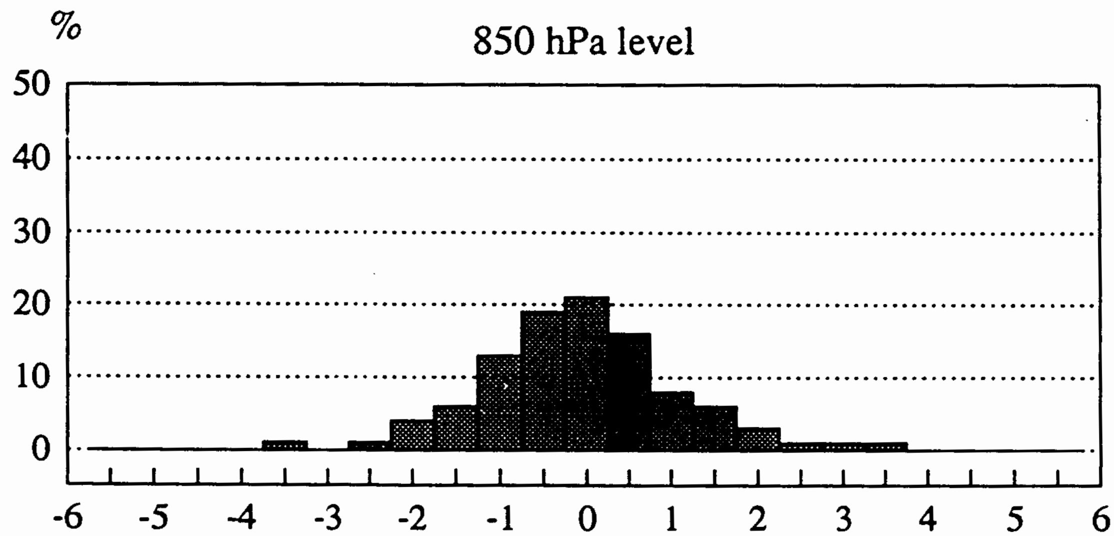


Fig. 12. Same as figure 4 except for normalized V component of the wind.

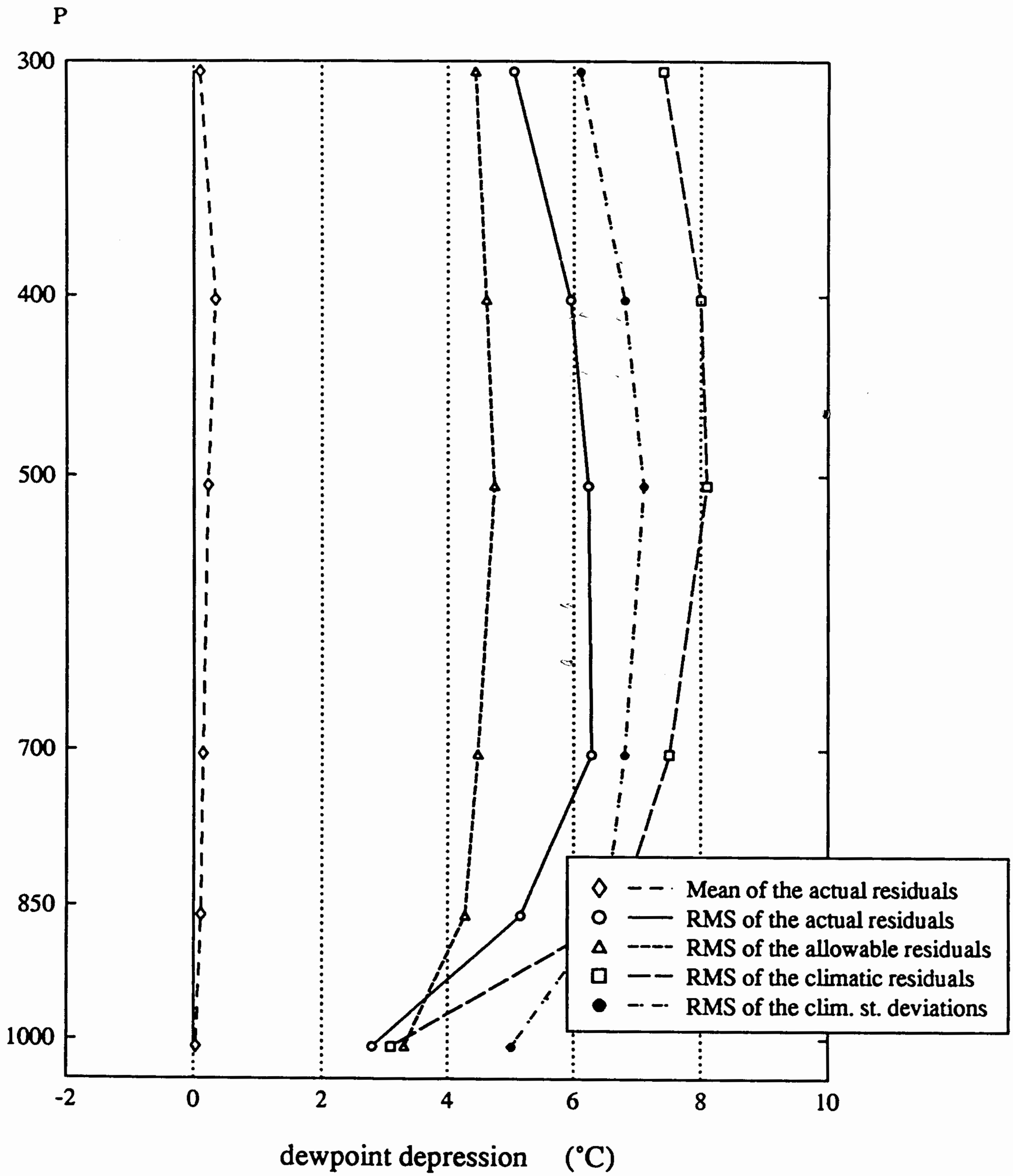


Fig. 13. Same as figure 3 except for dew point depression.

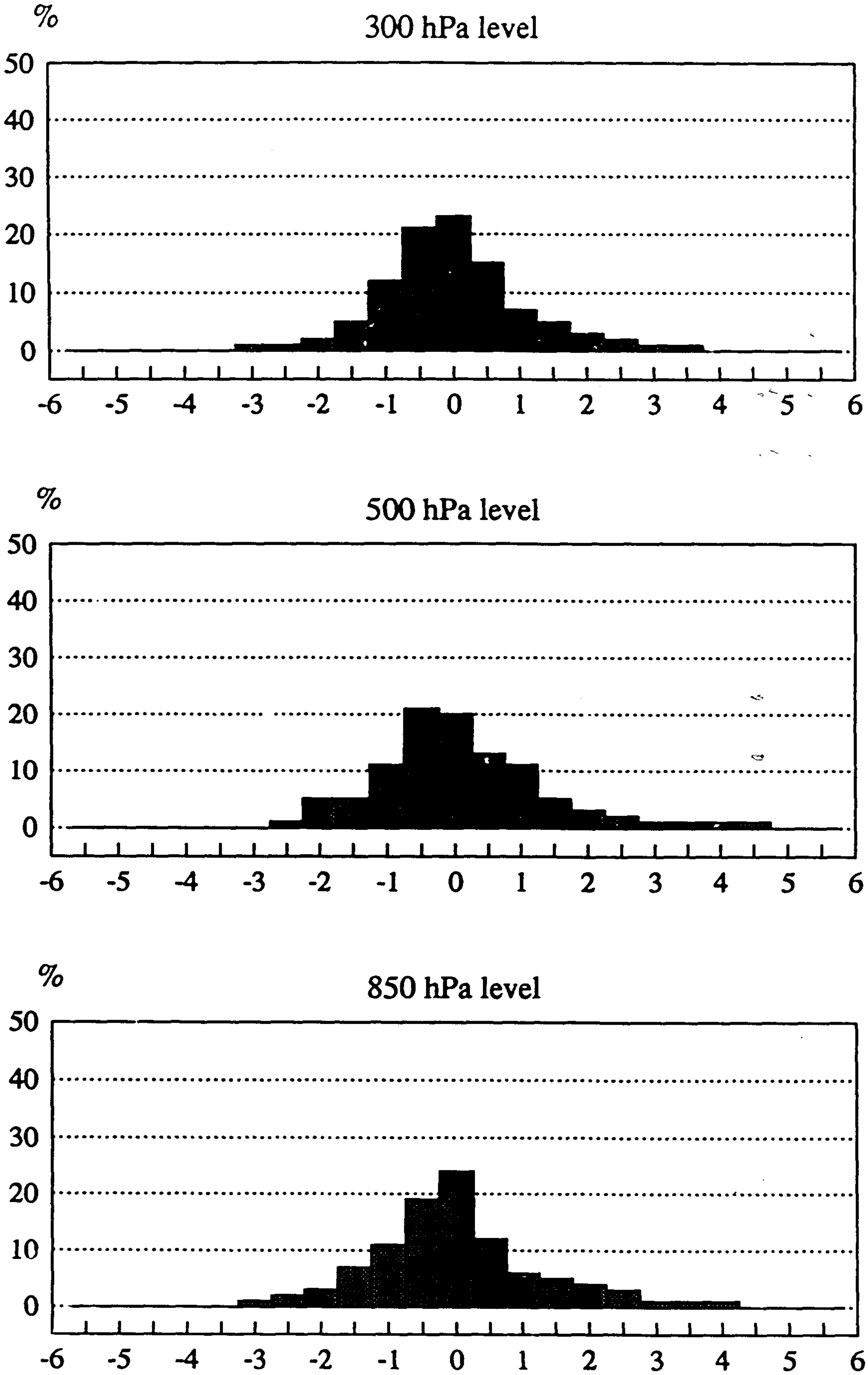


Fig. 14. Same as figure 4 except for dewpoint depression.

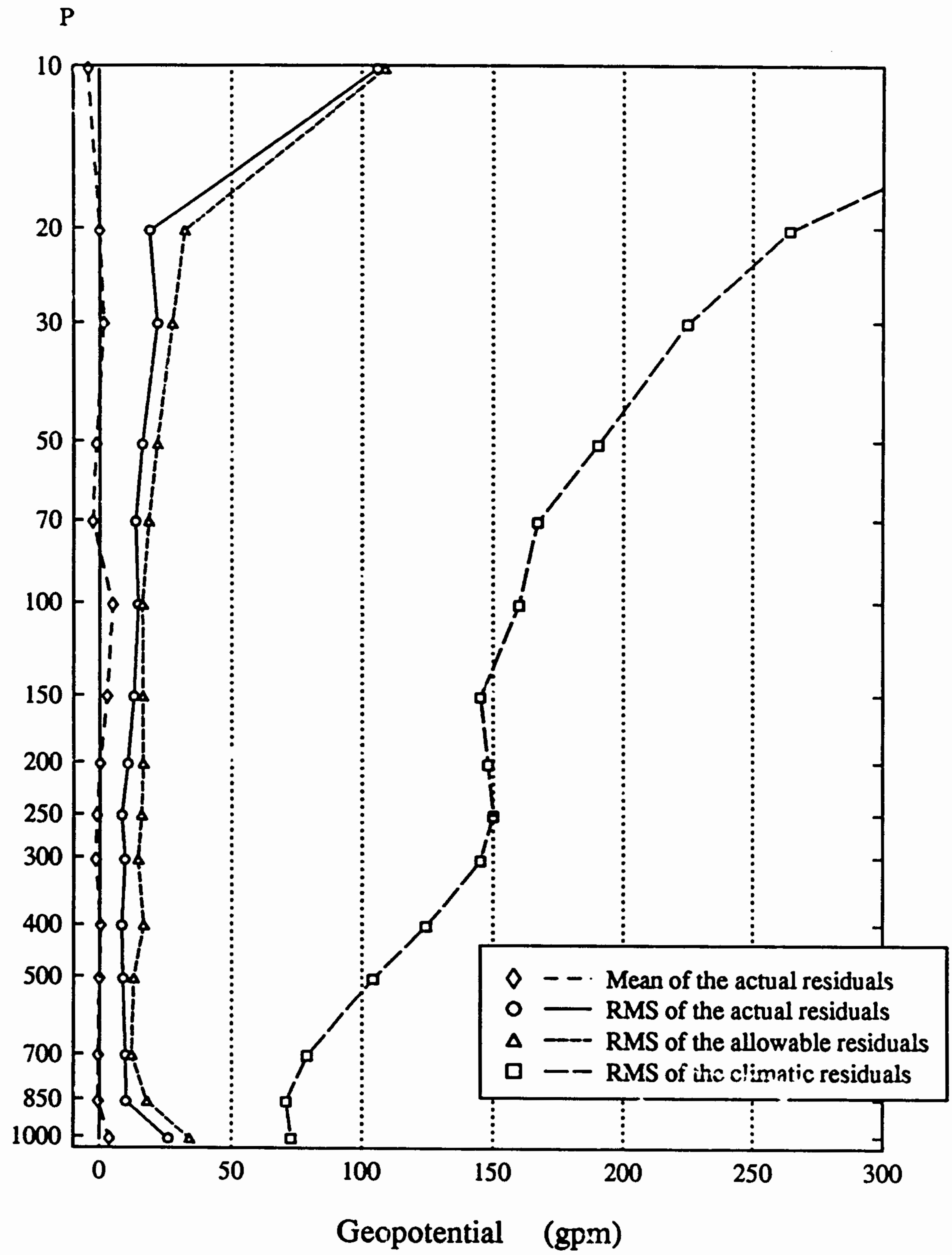


Fig. 15. Characteristics of vertical optimal interpolation of geopotential height for a data set of 759 stations from 00 UTC, 15 Jan 1989.

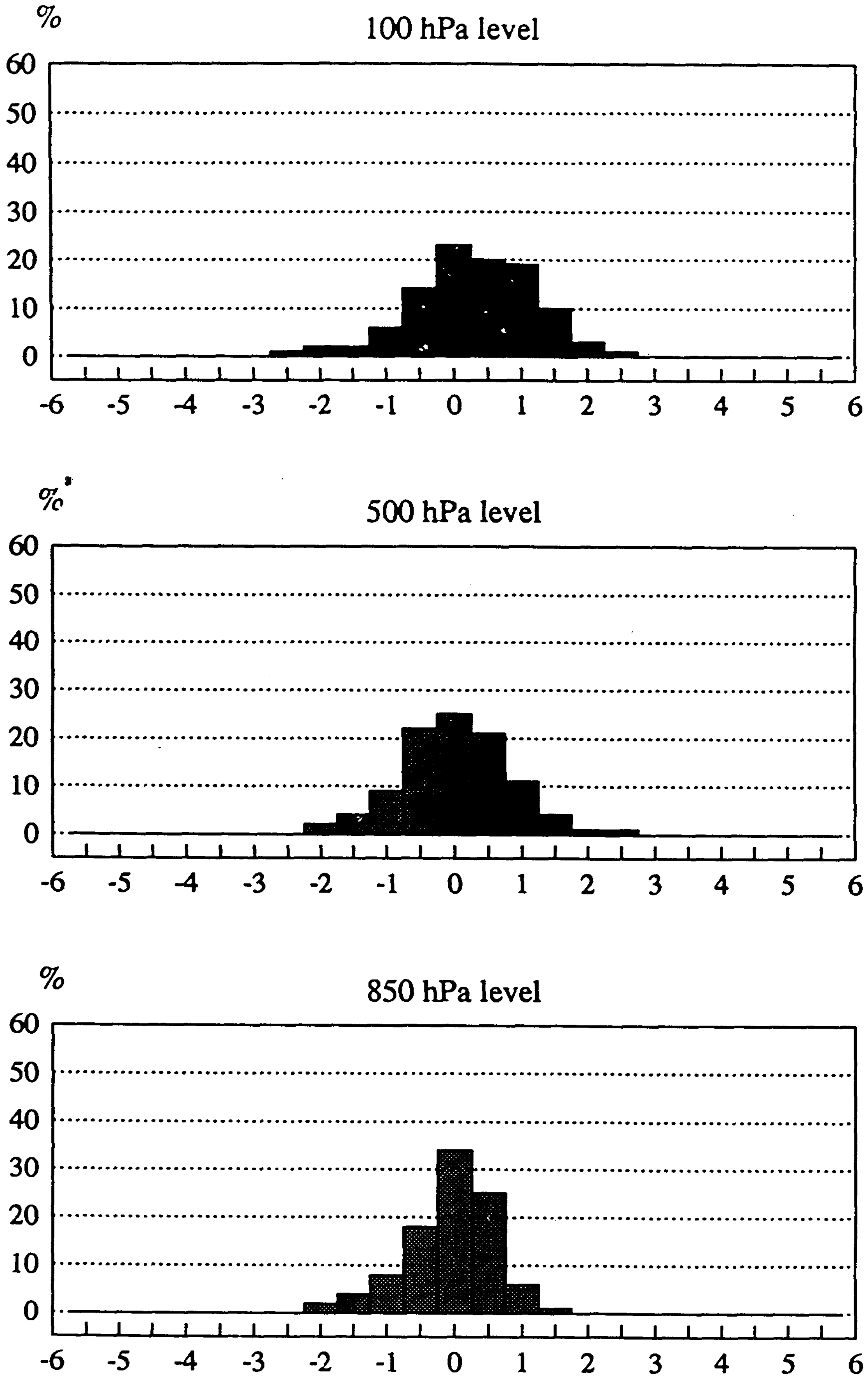


Fig. 16. Distribution of the normalized actual residuals from vertical optimal interpolation of geopotential height for a dataset of 759 station from 00 UTC, 15 Jan 1989.

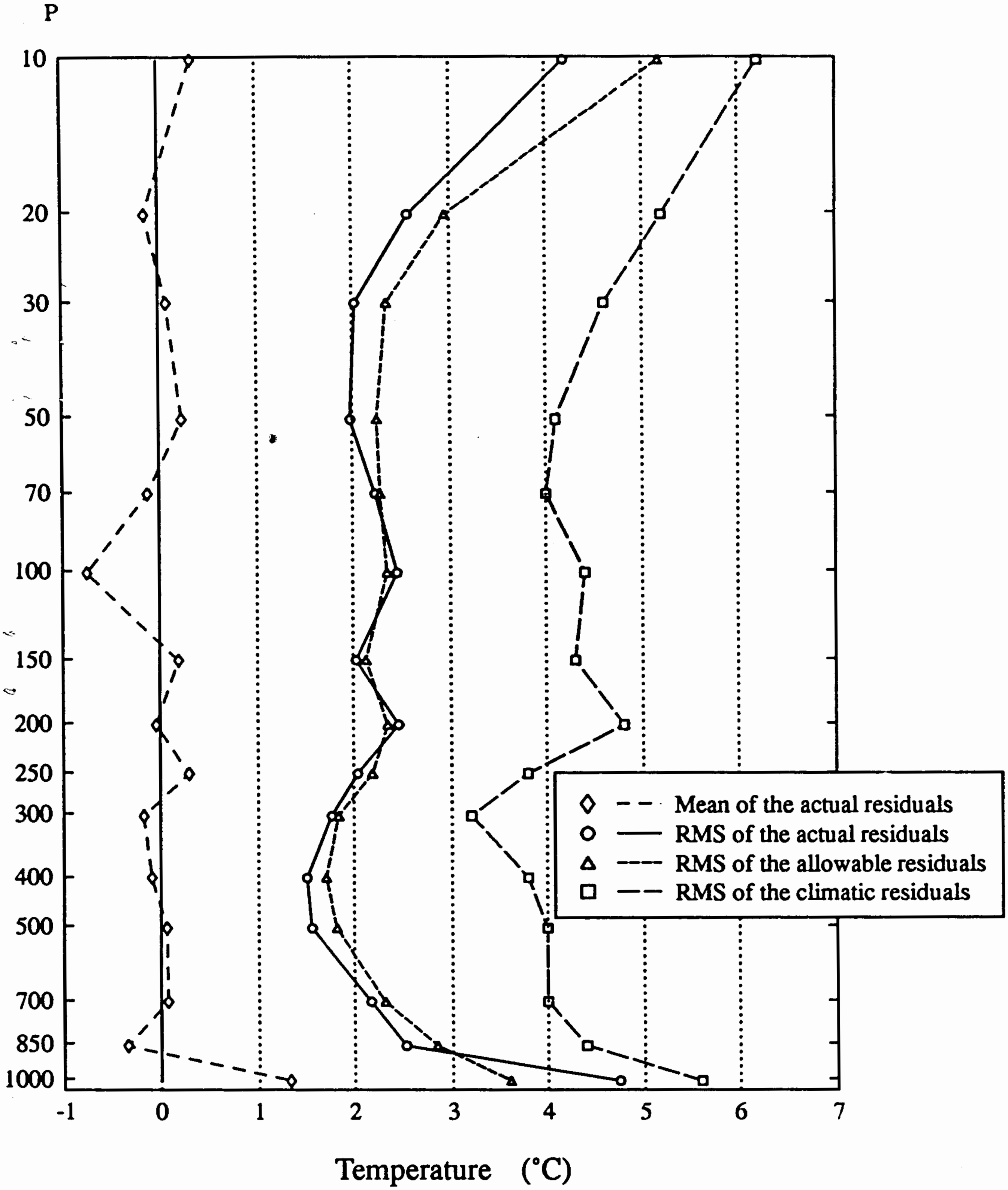


Fig. 17. Same as figure 17 except for temperature.

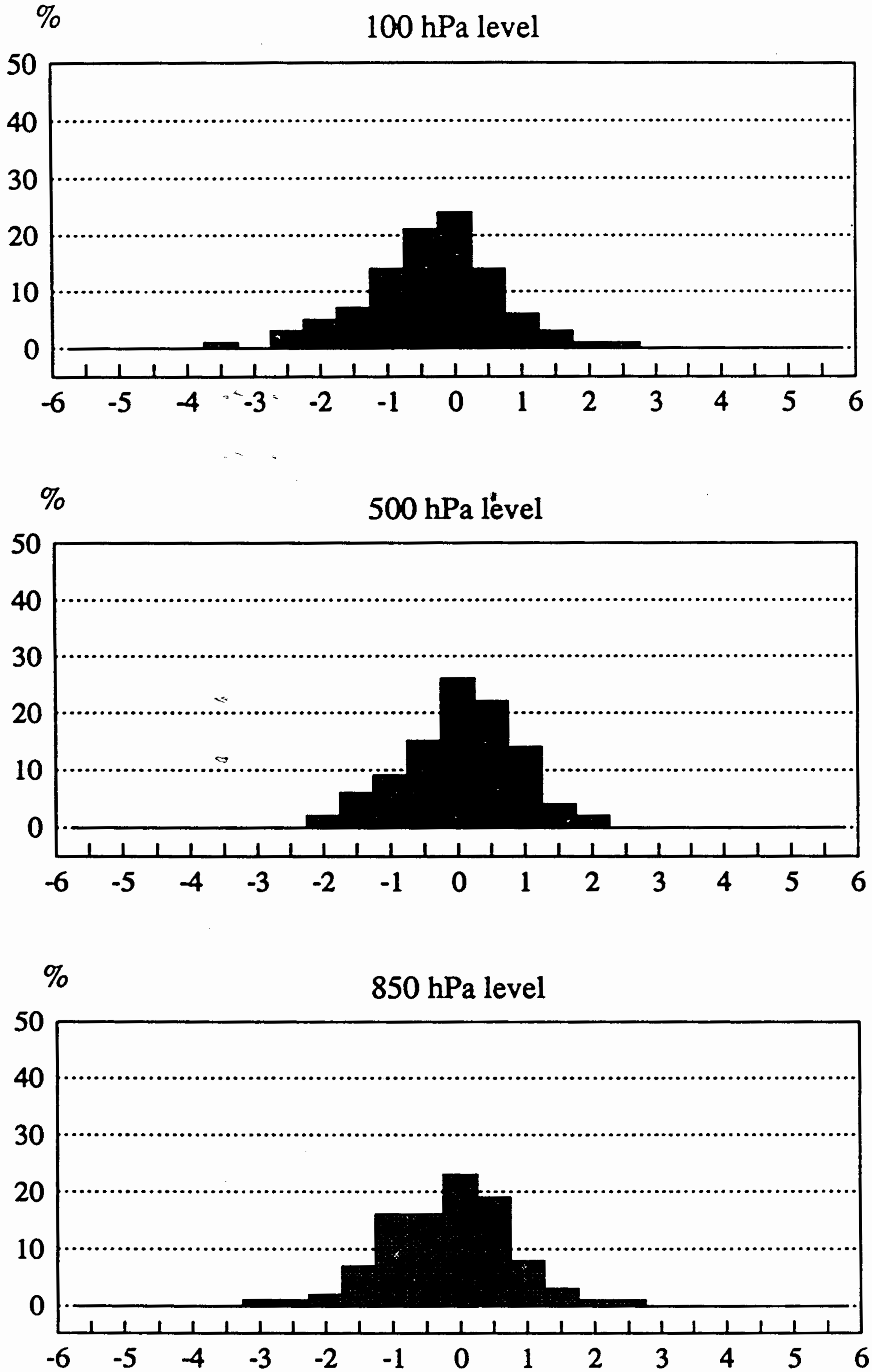


Fig. 18. Same as figure 16 except for temperature.

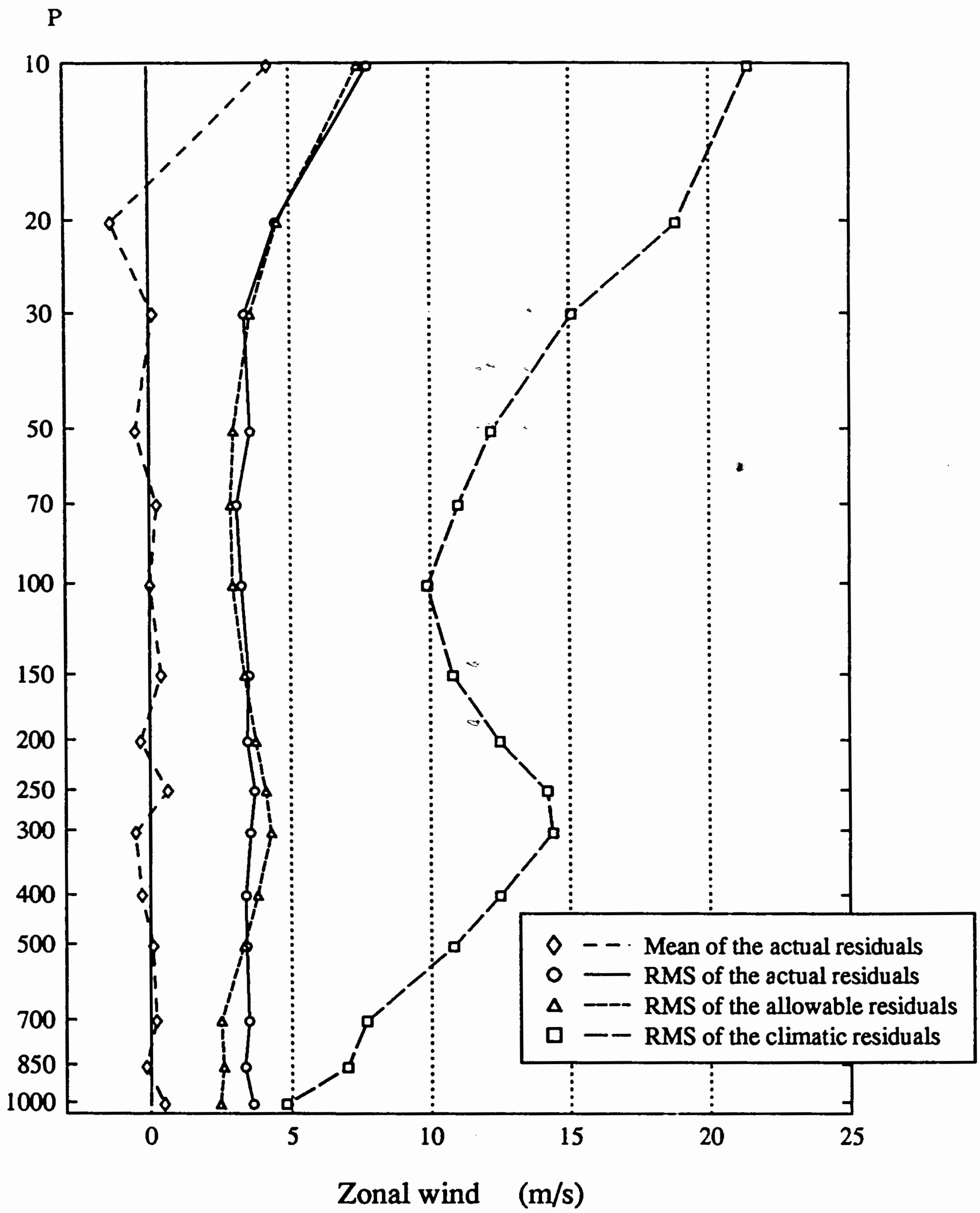


Fig. 19. Same as figure 15 except for the zonal wind component U.

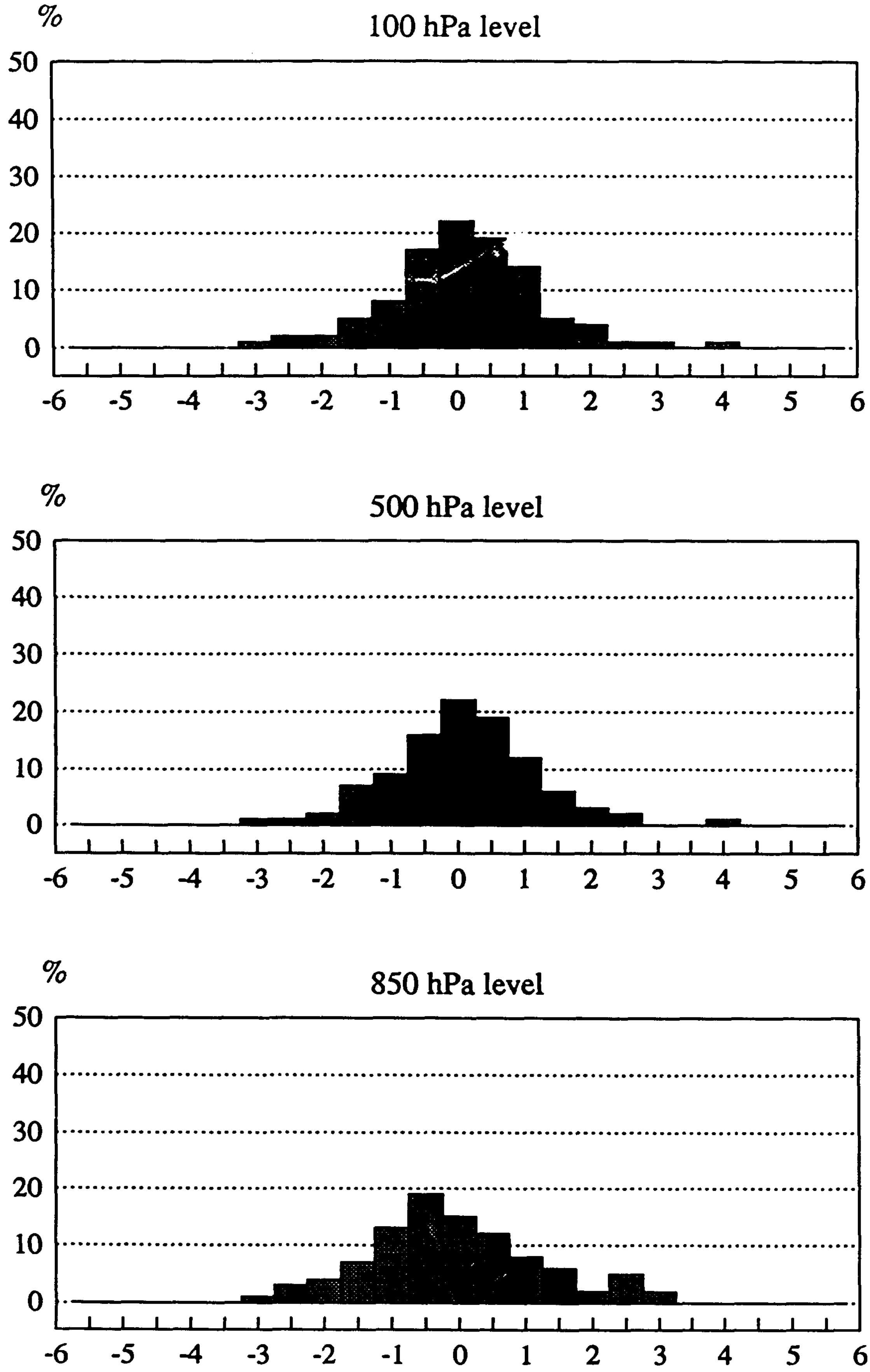


Fig. 20. Same as figure 16 except for zonal wind component, U.

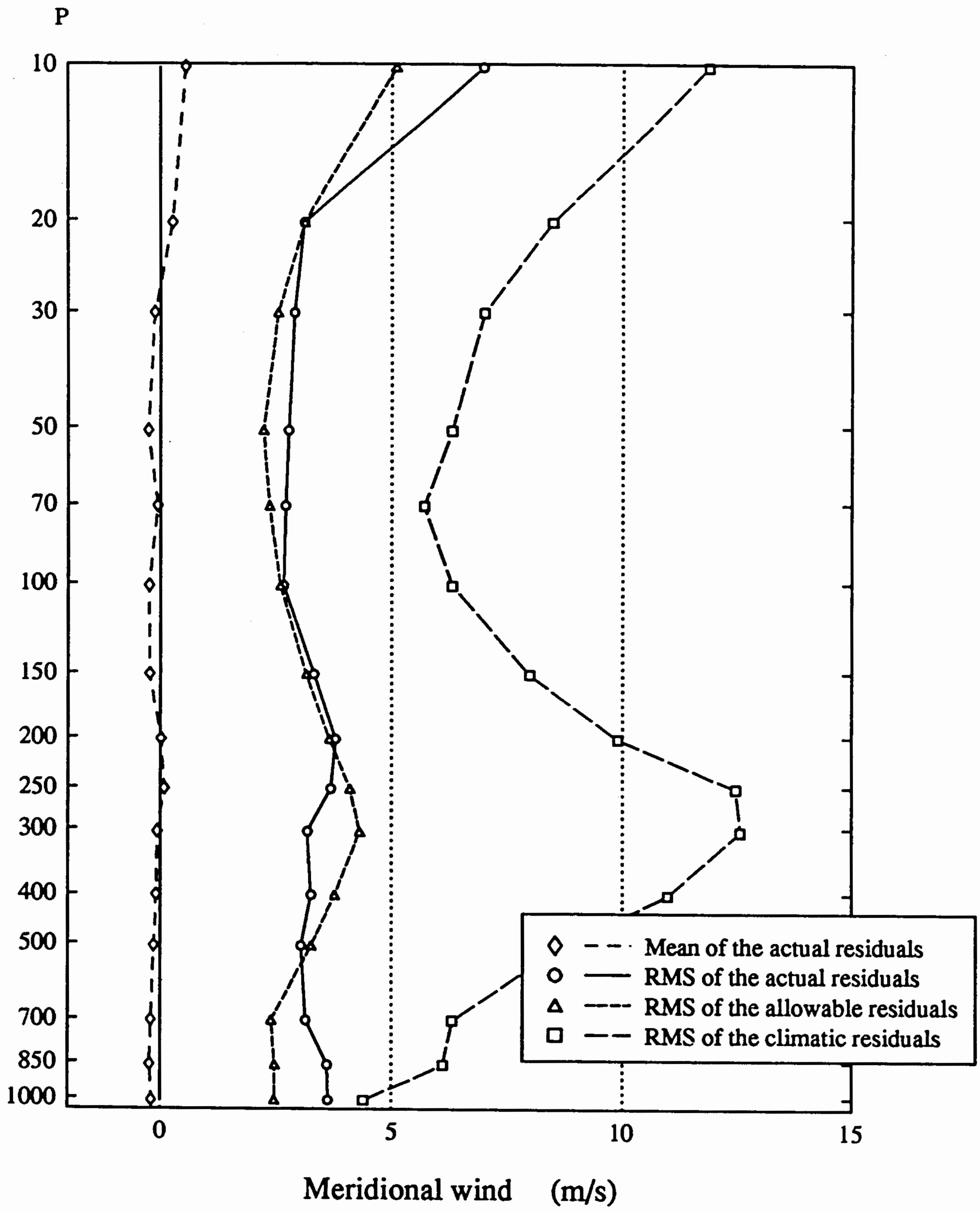


Fig. 21. Same as figure 15 except for the meridonal component of the wind, V.

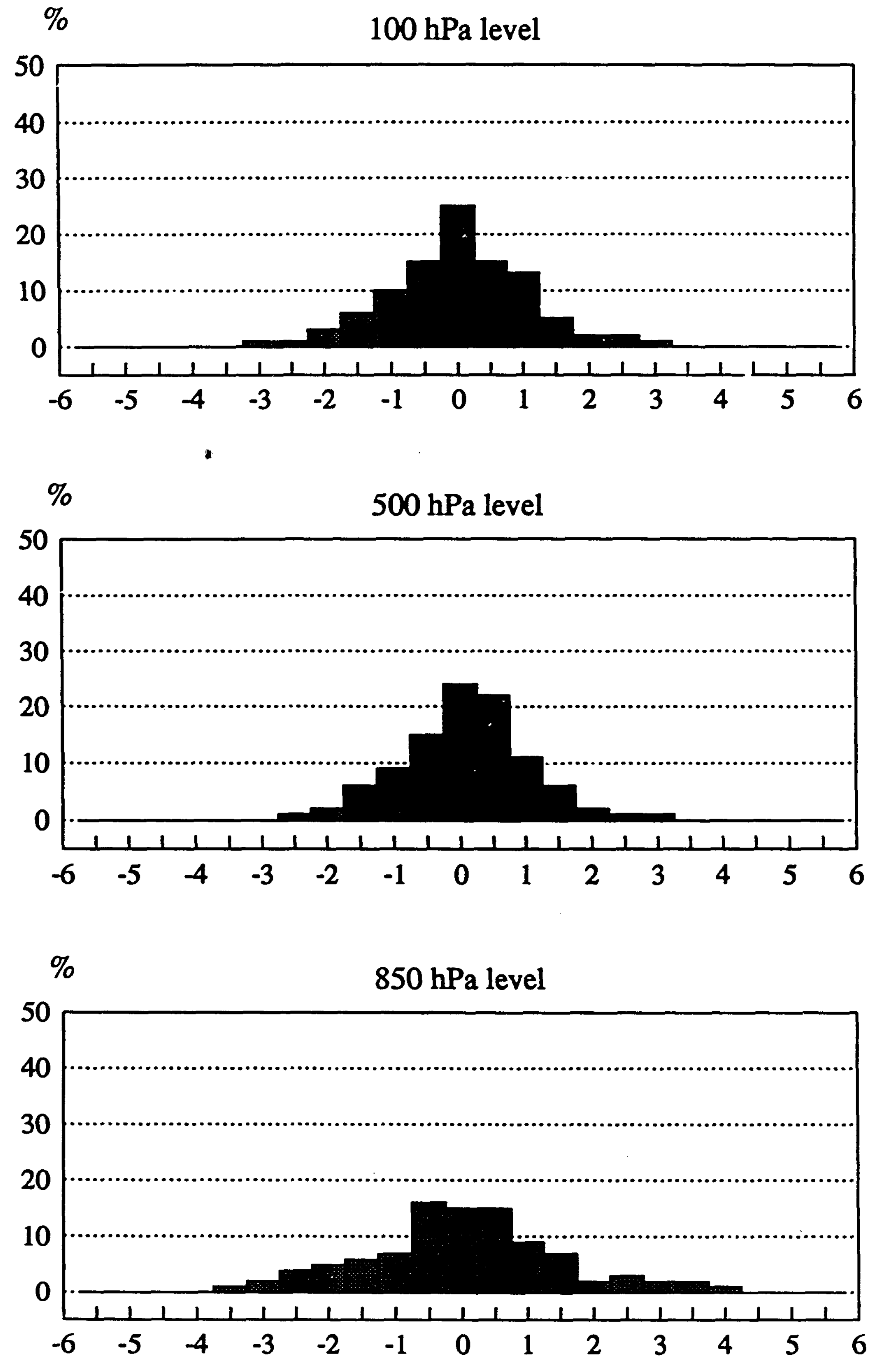


Fig. 22. Same as figure 16 except for meridional wind component, V

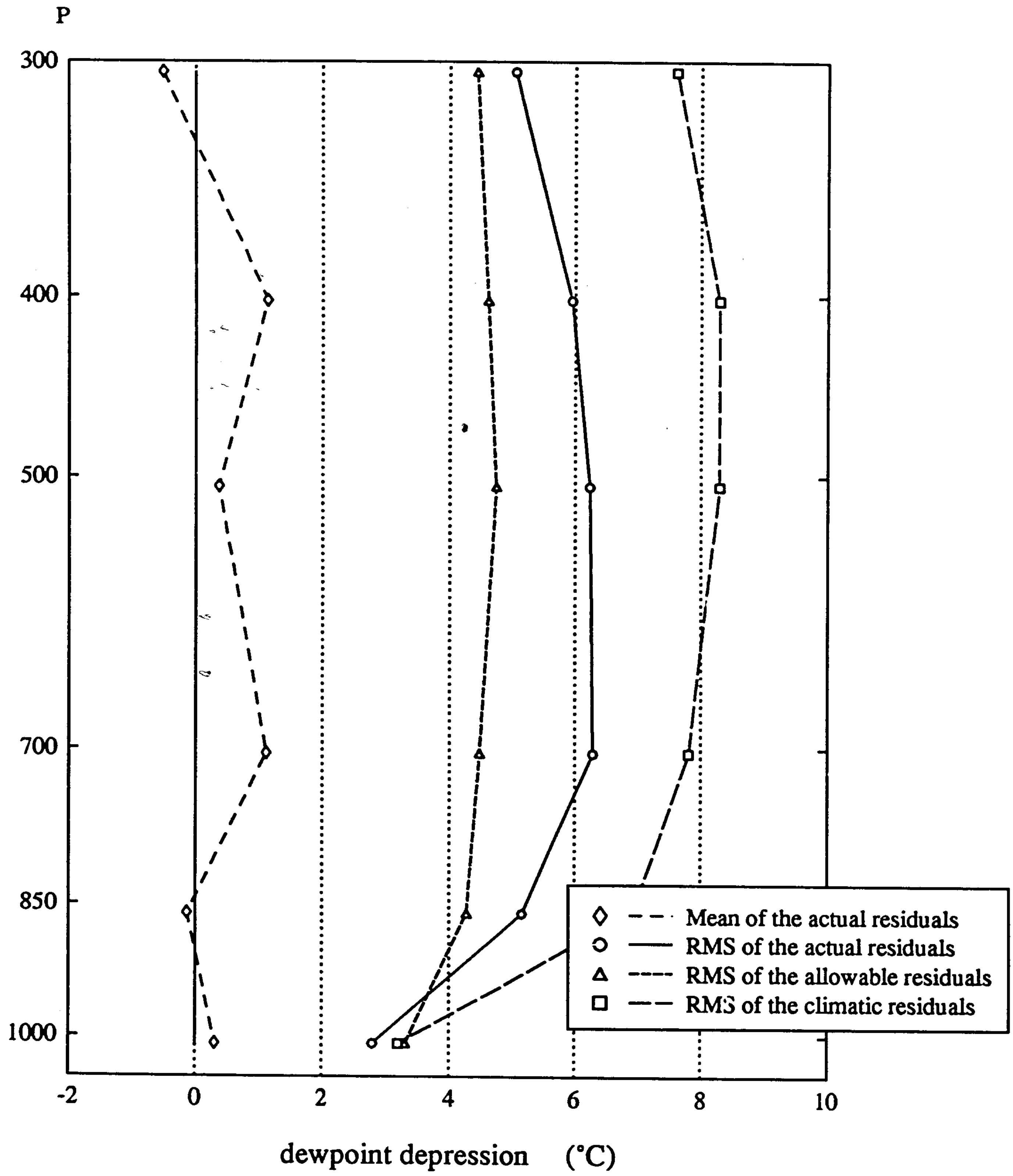


Fig. 23. Same as figure 15 except for the of dewpoint depression.

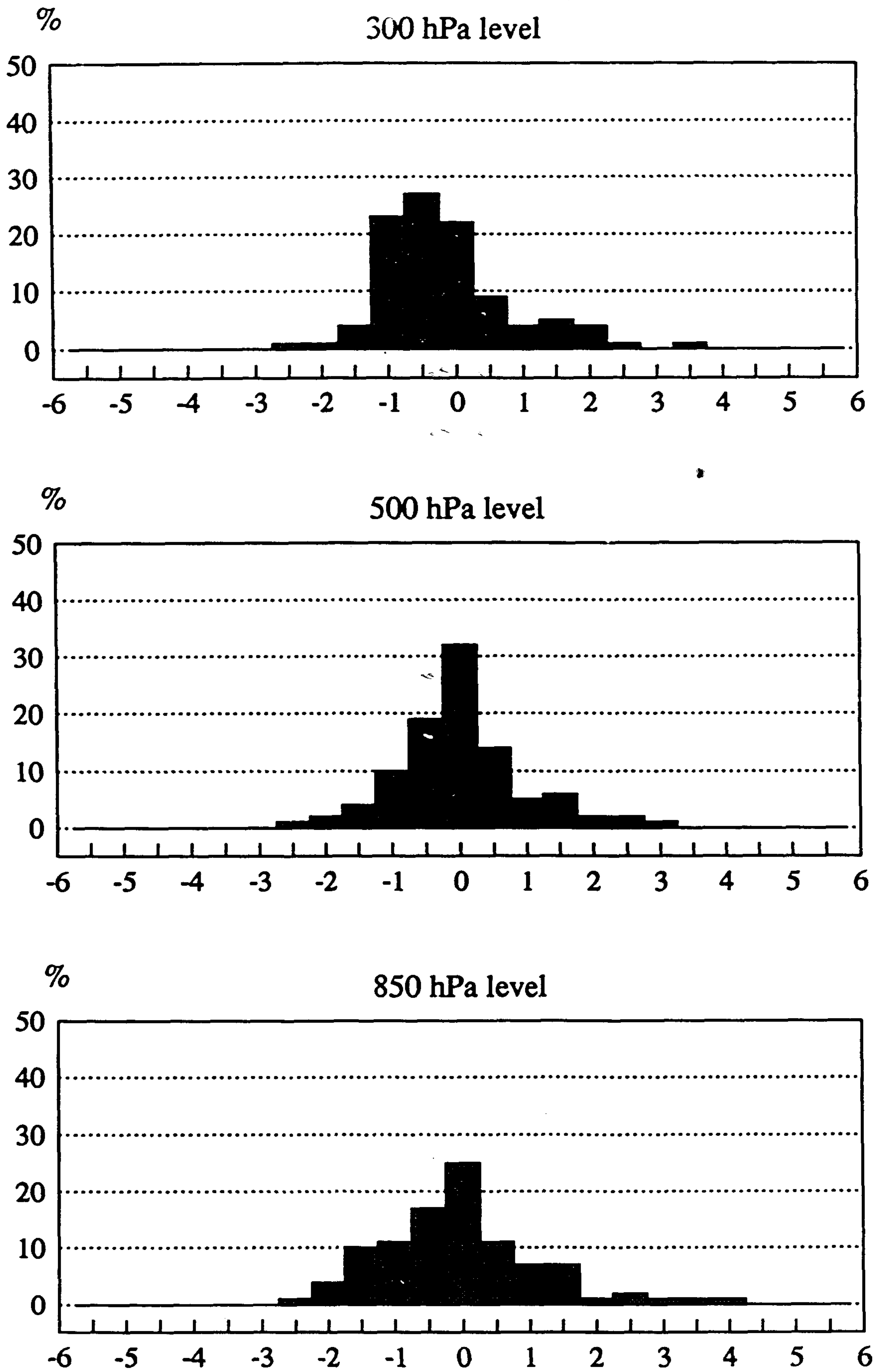


Fig. 24. Same as figure 16 except for dewpoint depression.

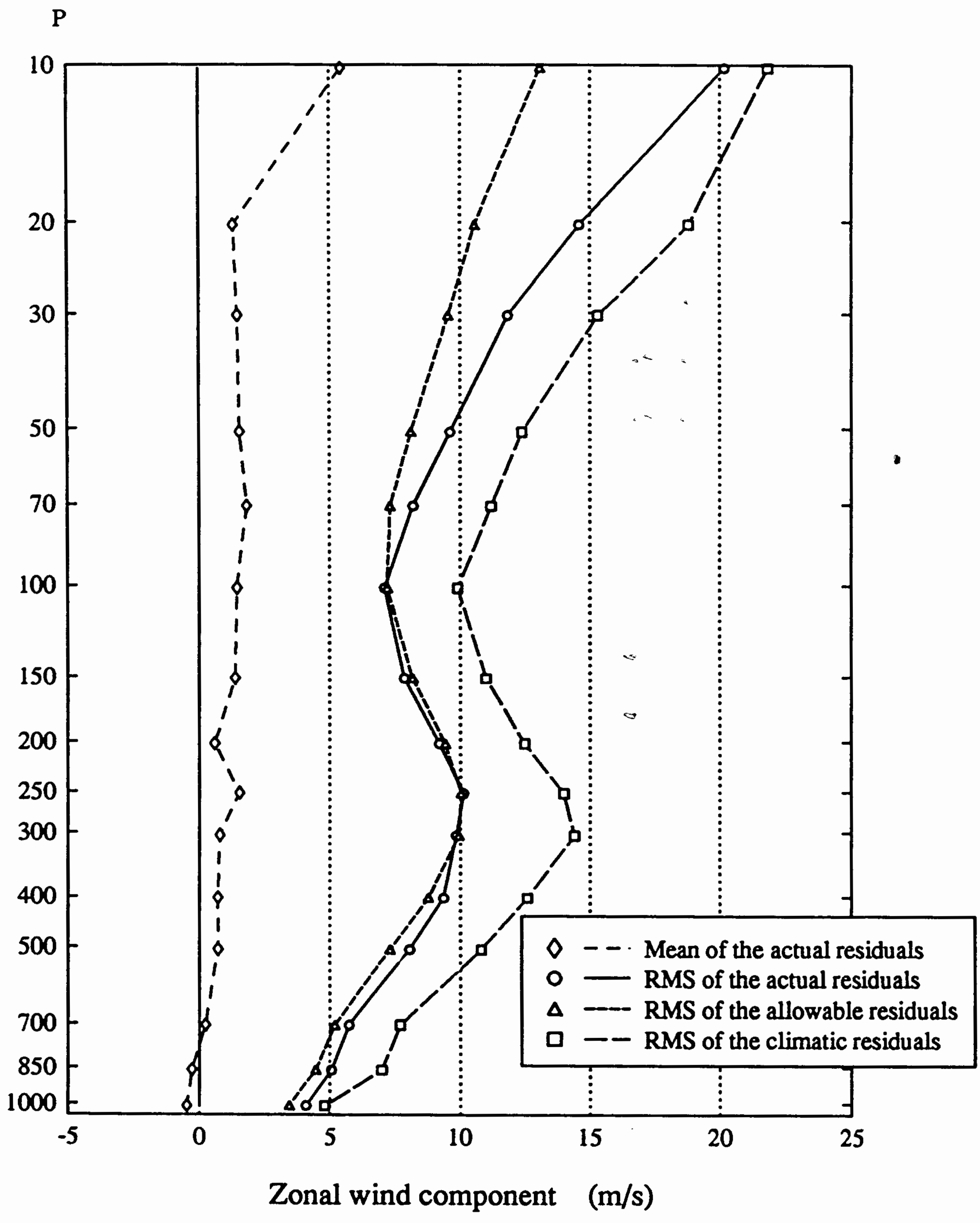


Fig. 25. Characteristics of geostrophic approximation to the zonal wind component for a dataset of 759 stations from 00 UTC, 15 Jan 1989.

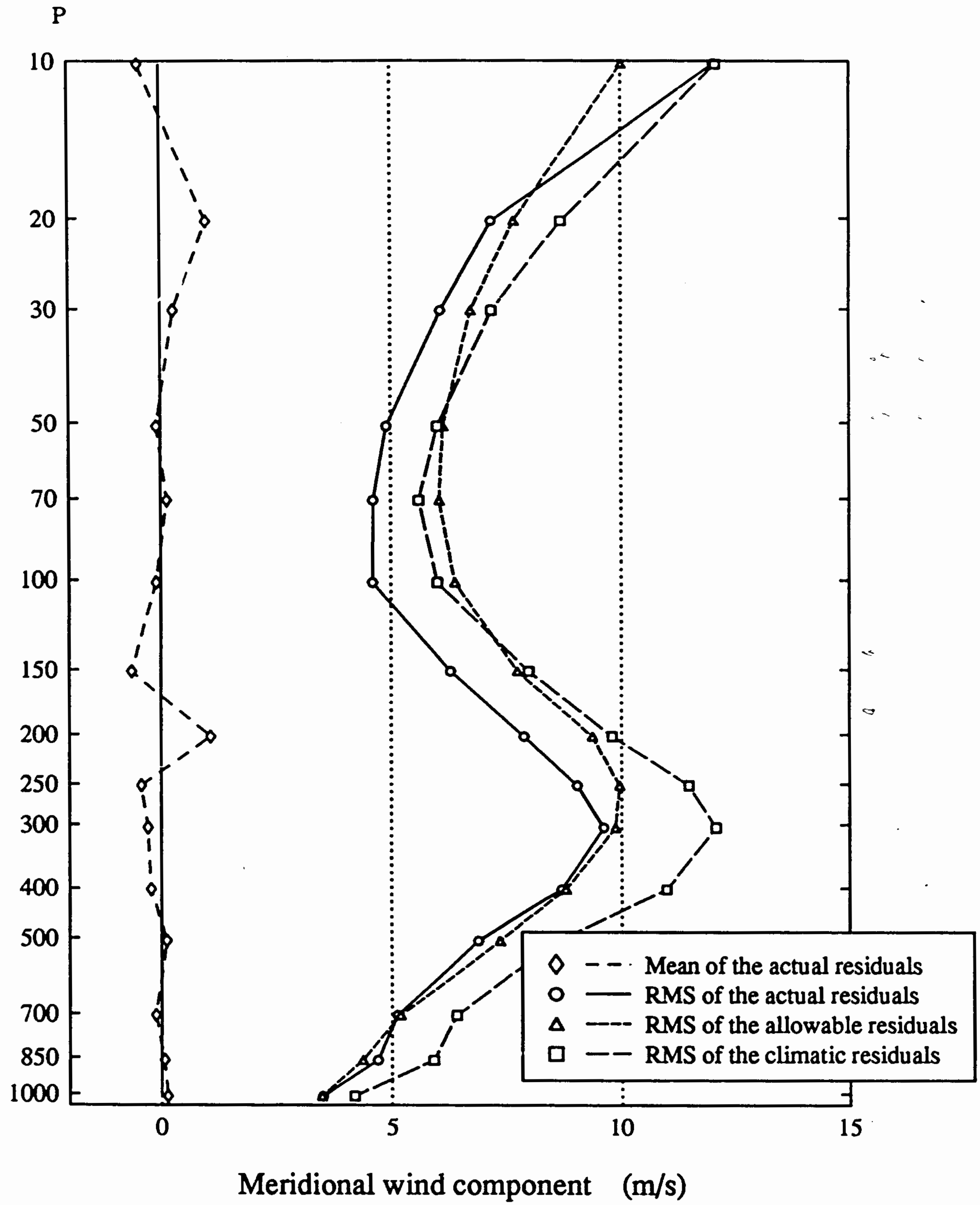


Fig. 26. Same as figure 25 except for the meridional wind component.

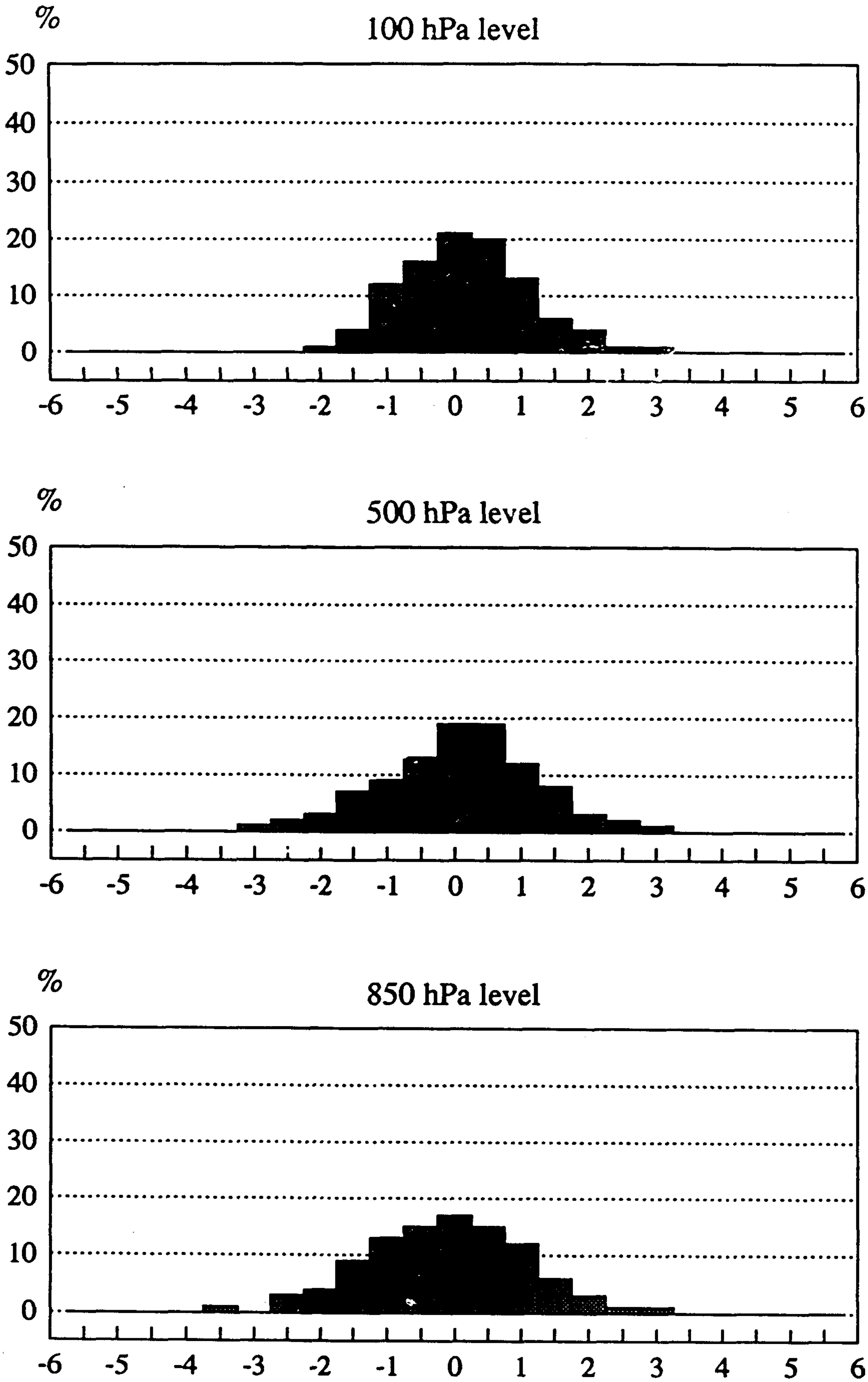


Fig. 27. Distribution of the normalized residuals from the geostrophic approximation of the U component of the wind for a global dataset of 759 stations from 00 UTC, 15 Jan 1989.

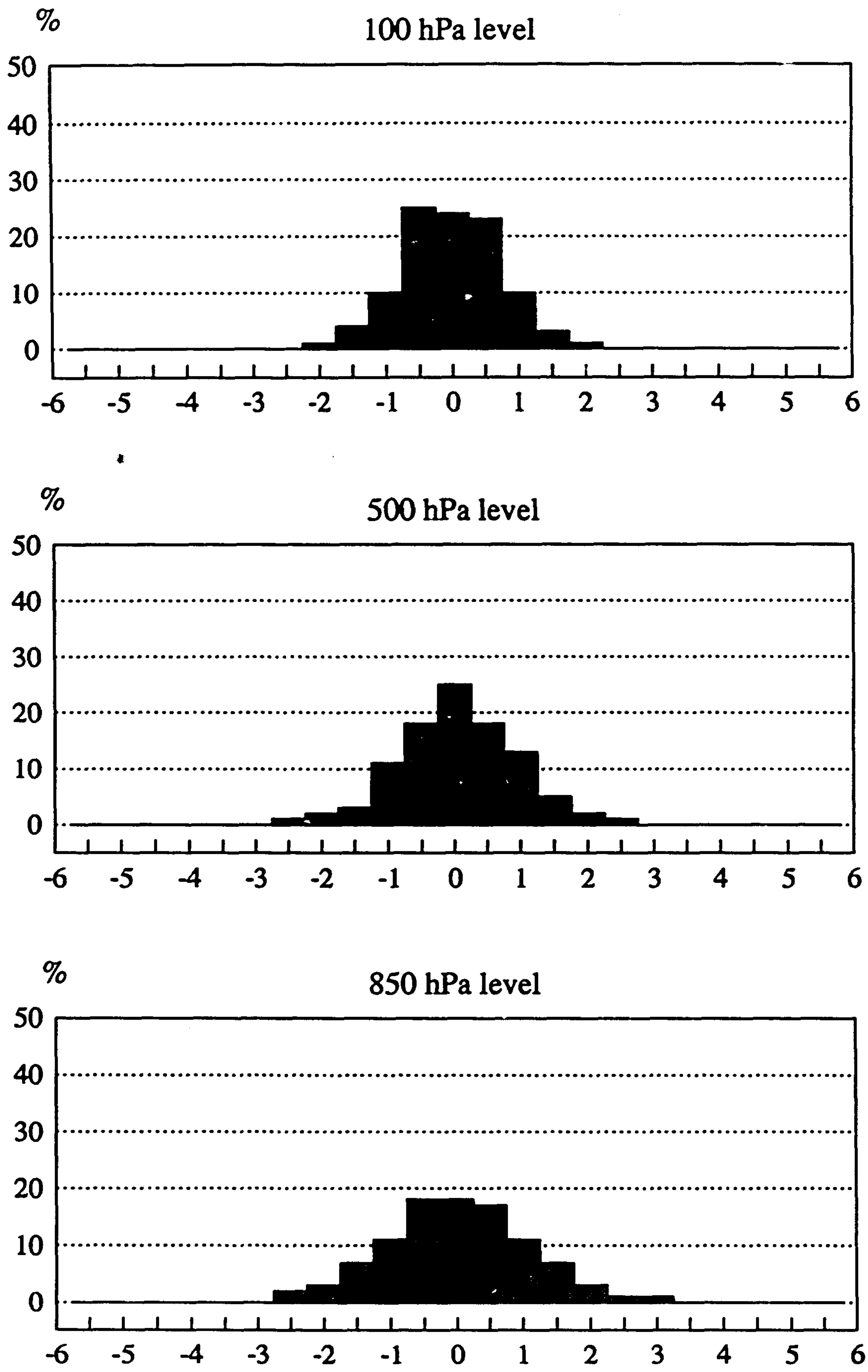


Fig. 28. Same as figure 27 except for the meridional component.

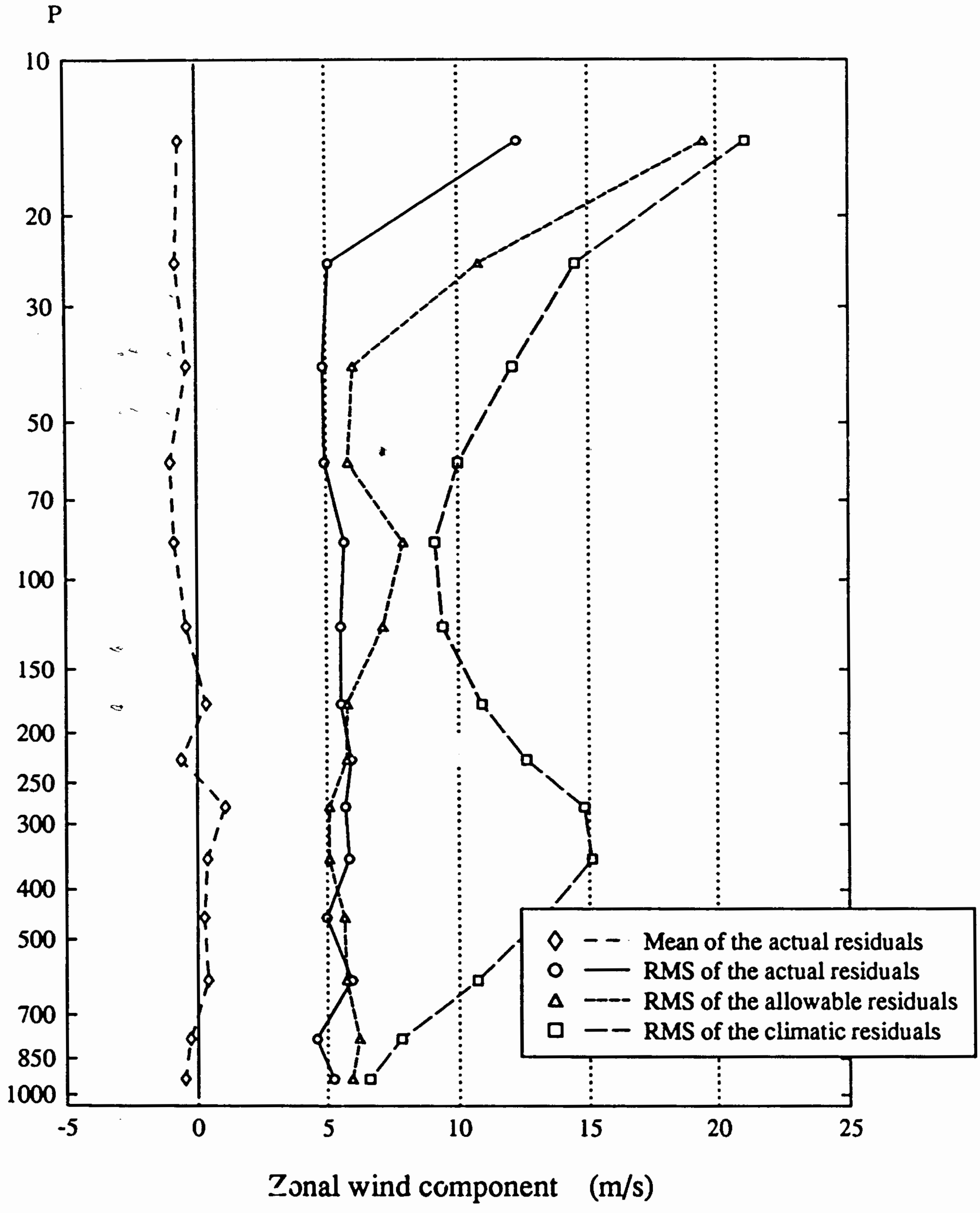


Fig. 29. Characteristics of the thermal wind approximation of the zonal wind shift between mandatory levels for a dataset of 759 stations from 00 UTC, 15 Jan 1989.

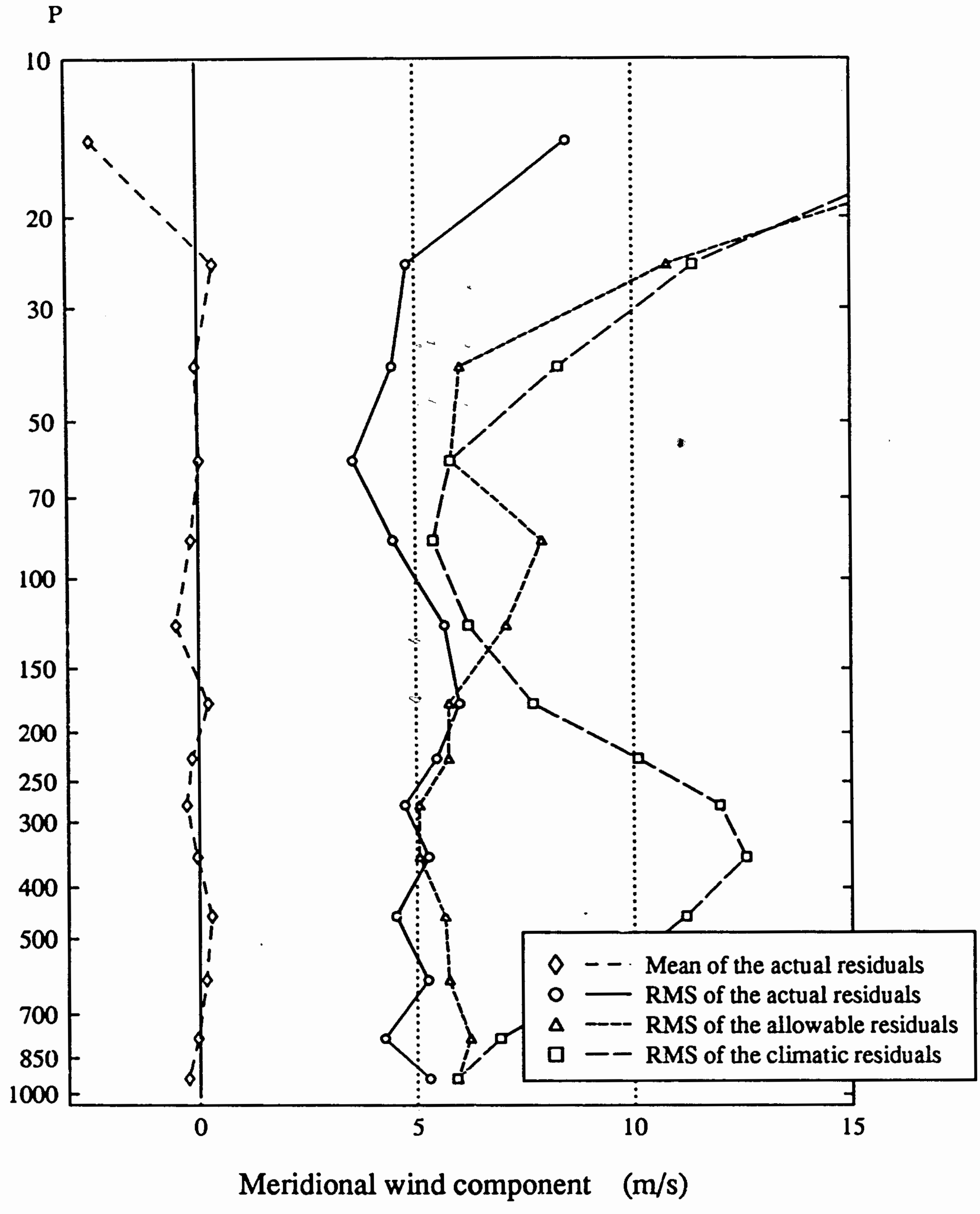


Fig. 30. Same as figure 29 except for the meridional wind component.

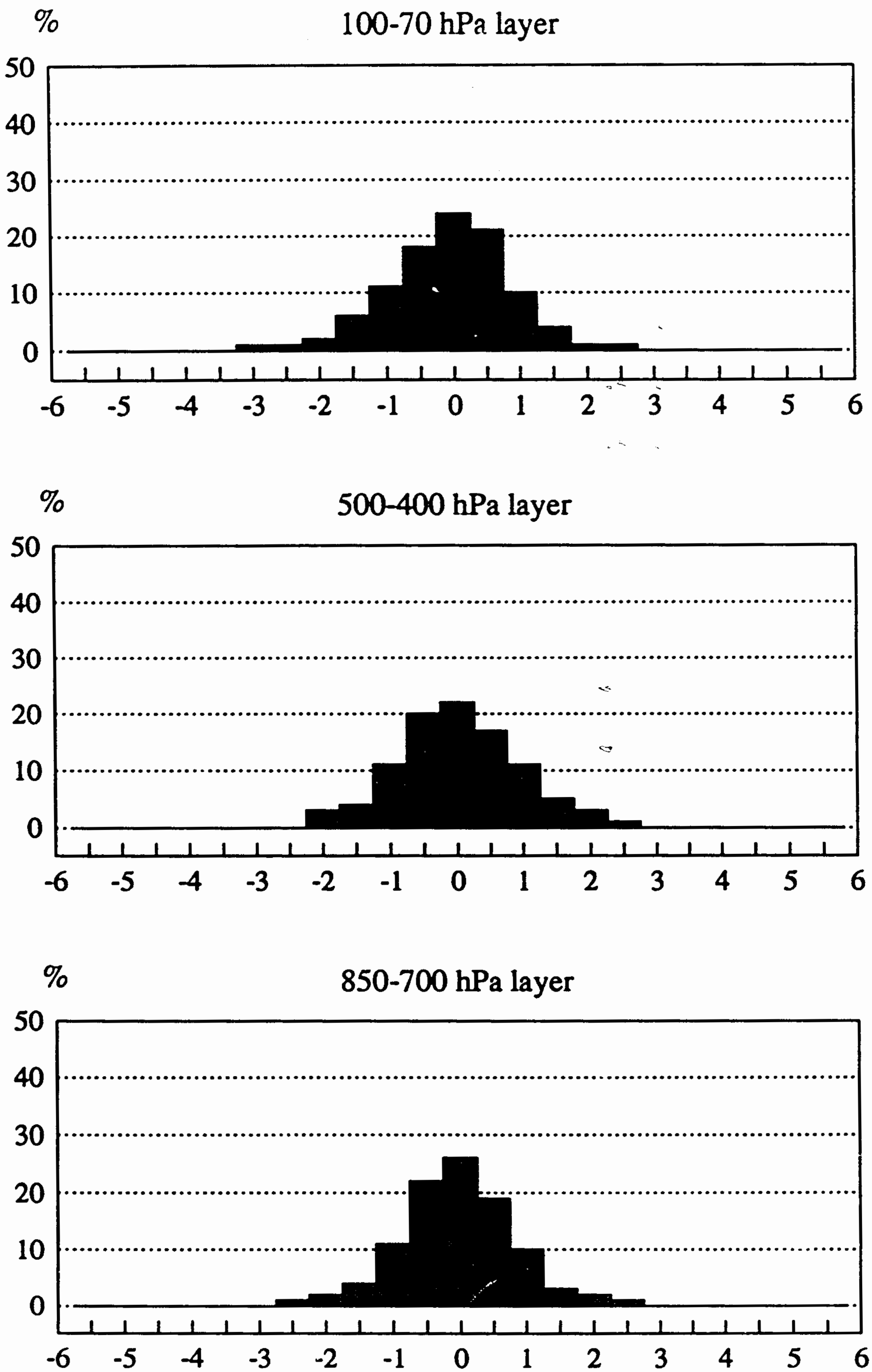


Fig. 31. Distribution of the normalized actual residuals for the thermal wind approximation of U component wind shift for a dataset of 759 stations from 00 UTC, 15 Jan 1989.

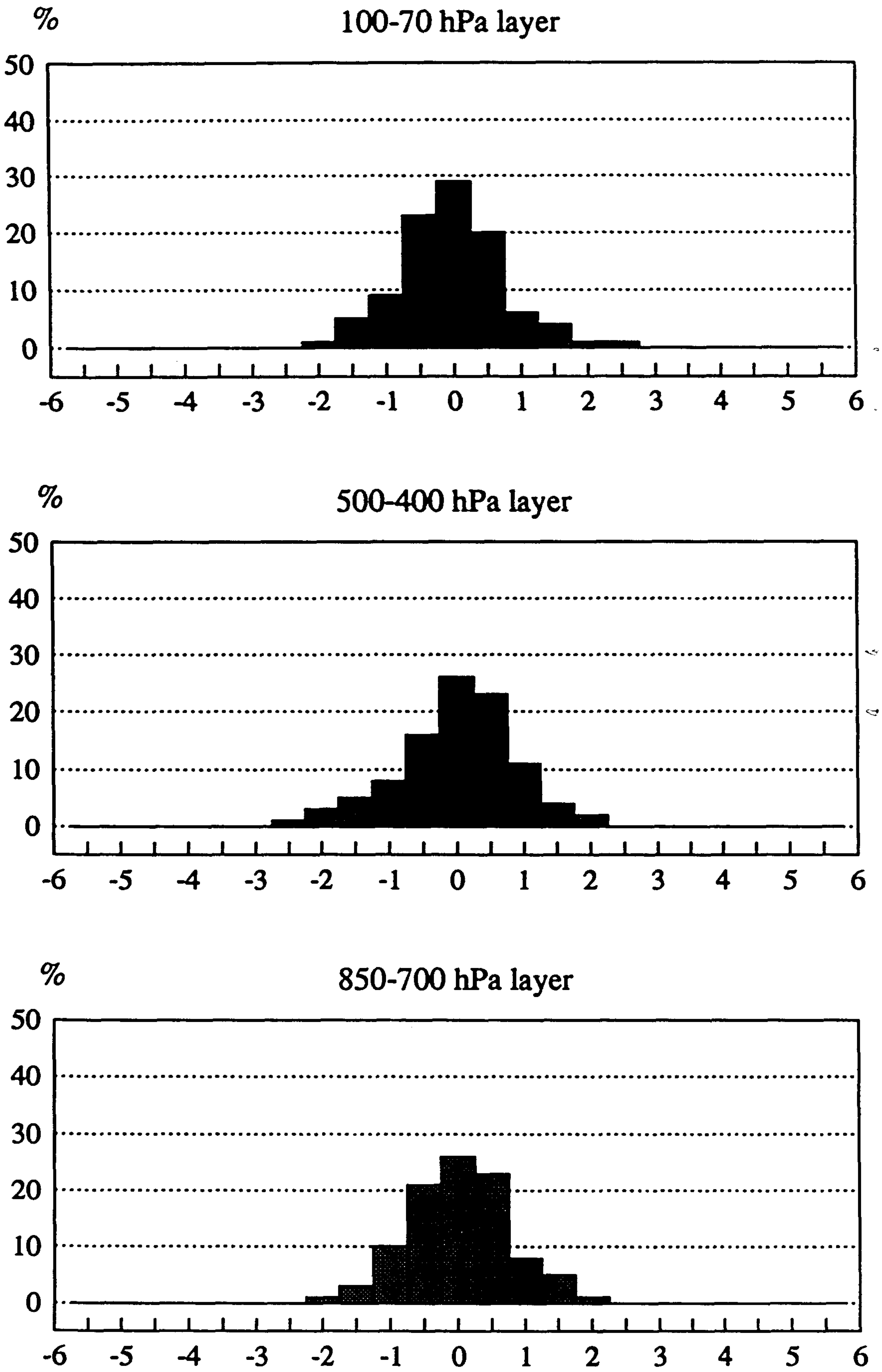


Fig. 32. Same as figure 31 except for the meridional wind component.

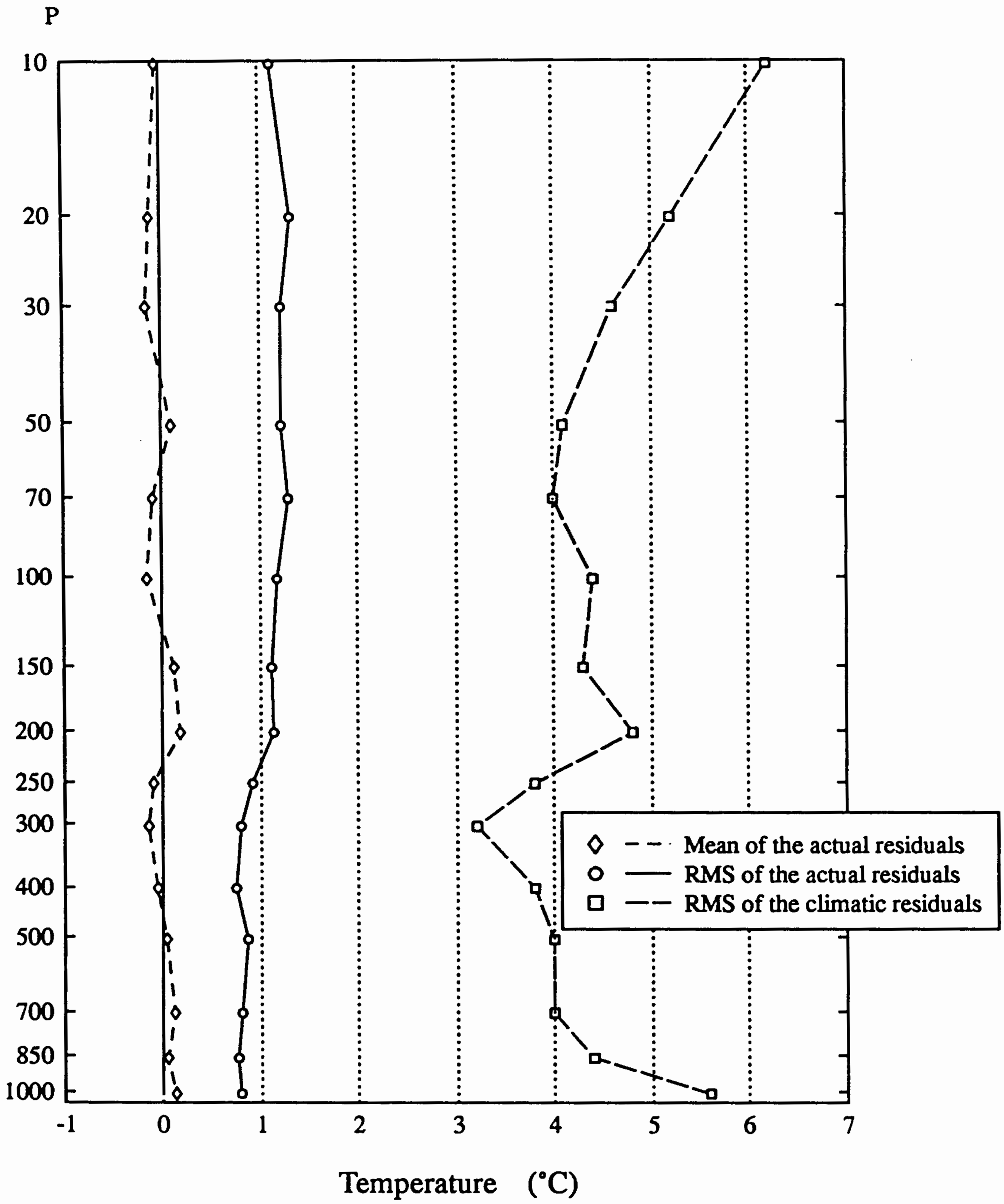


Fig. 33. Characteristics of linear interpolation of temperature from significant levels to mandatory levels for a dataset of 759 stations from 00 UTC, 15 Jan 1989.

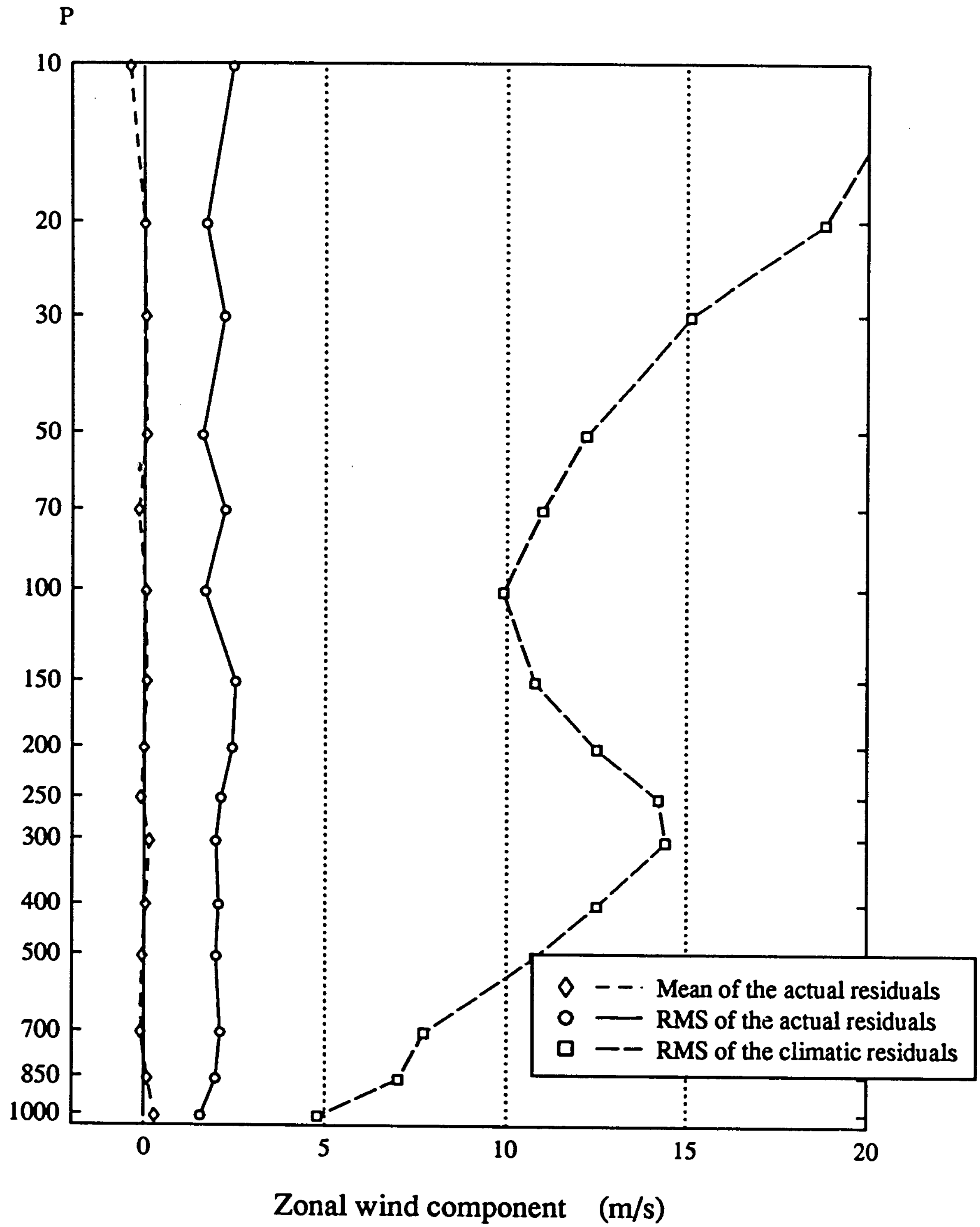


Fig. 34. Same as figure 33 except for the zonal wind component.

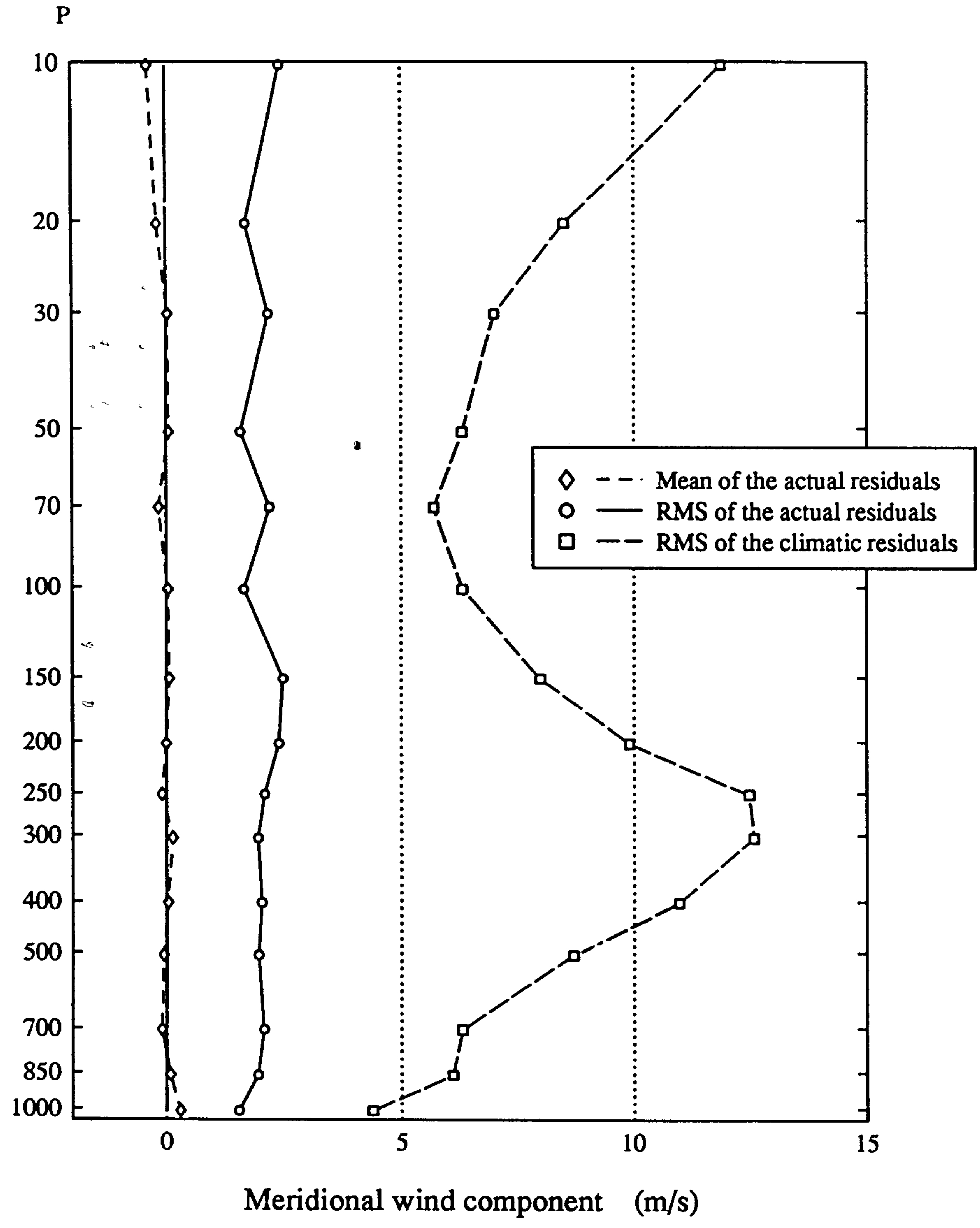


Fig. 35. Same as figure 33 except for the meridional wind component.

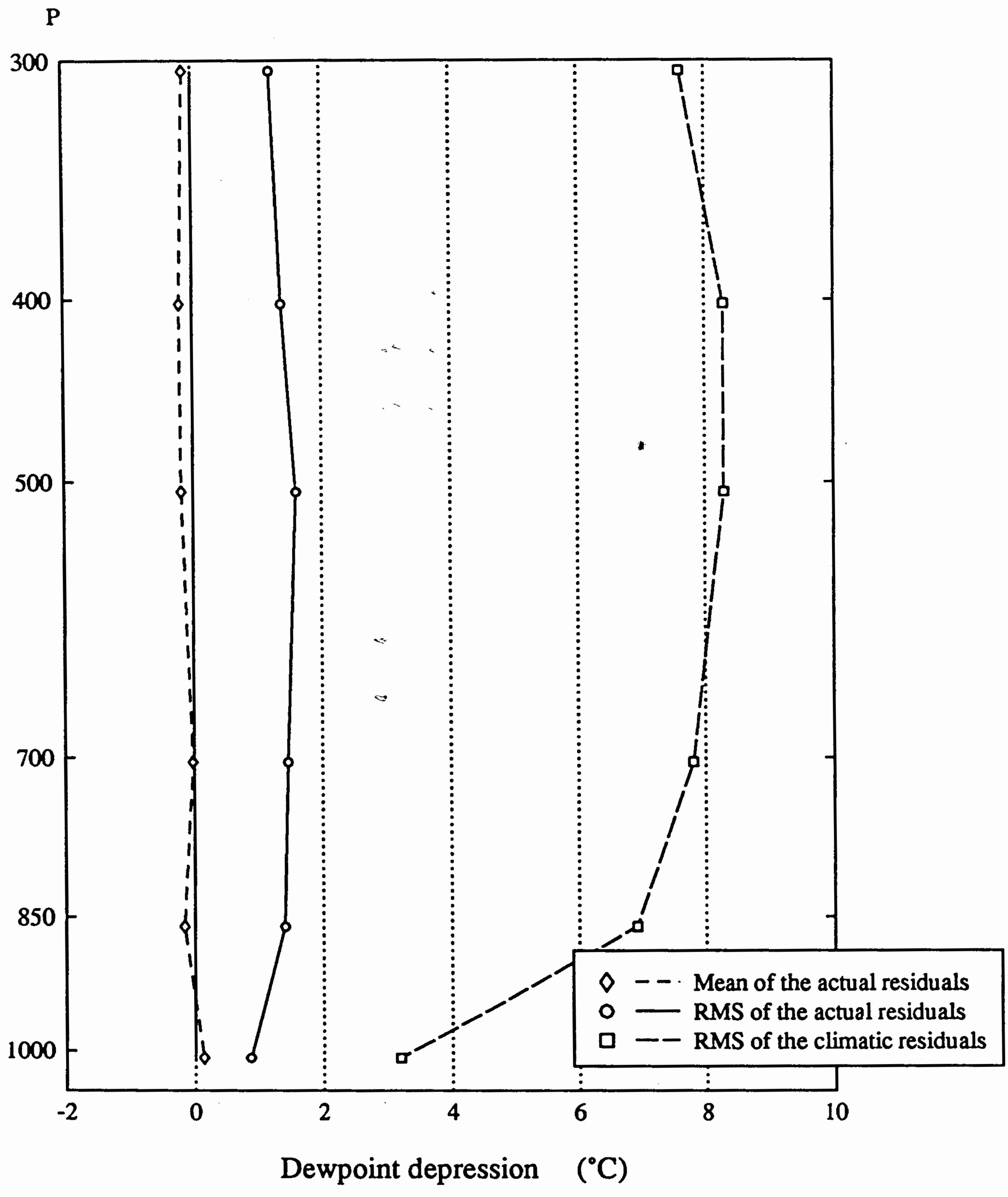


Fig. 36. Same as figure 33 except for dewpoint depression.

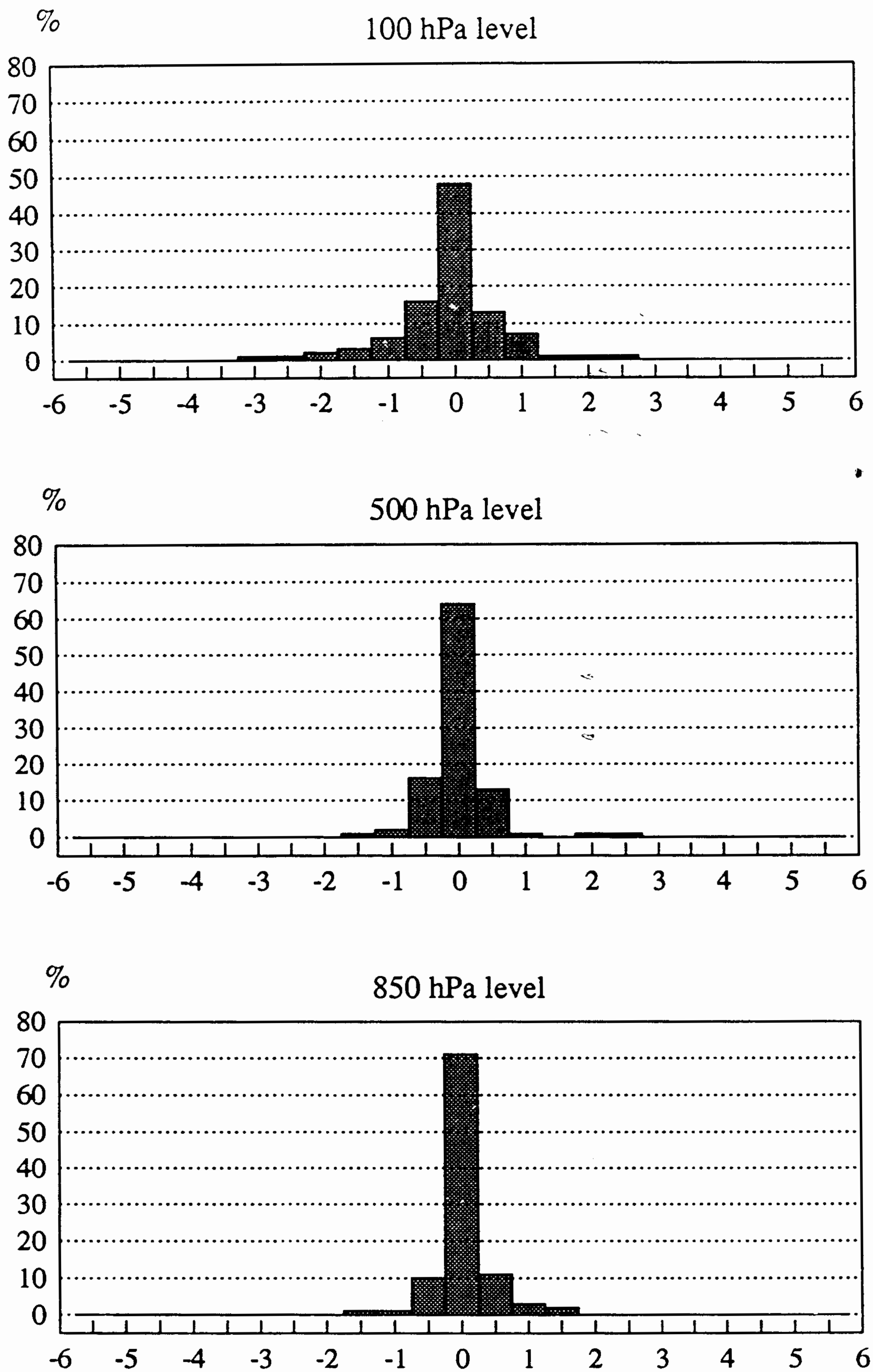


Fig. 37. Distribution of the normalized actual residuals from the linear interpolation of temperature from significant levels to mandatory levels for a dataset of 759 stations from 00 UTC, 15 Jan 1989.

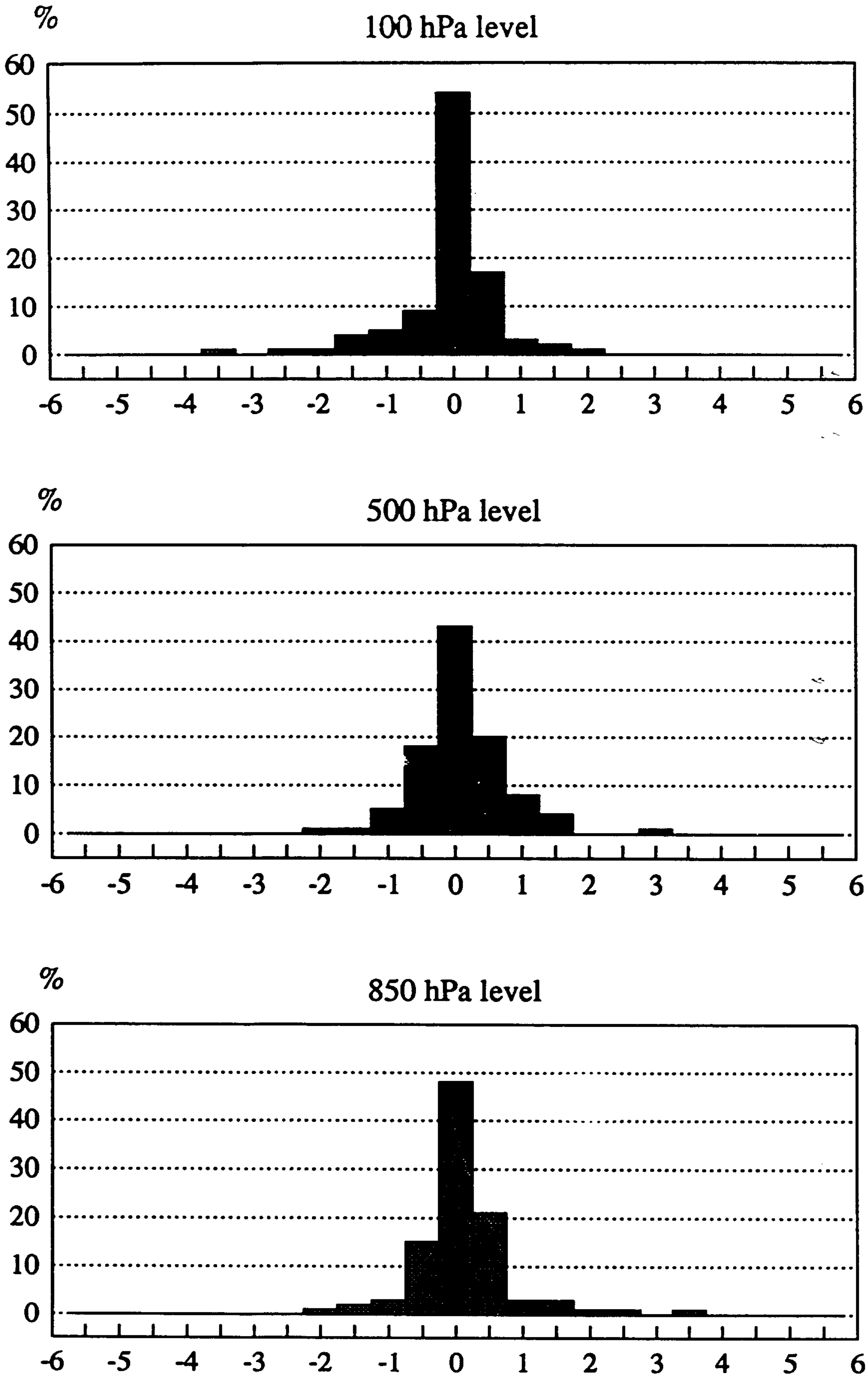


Fig. 38. Same as figure 37 except for the zonal wind component.

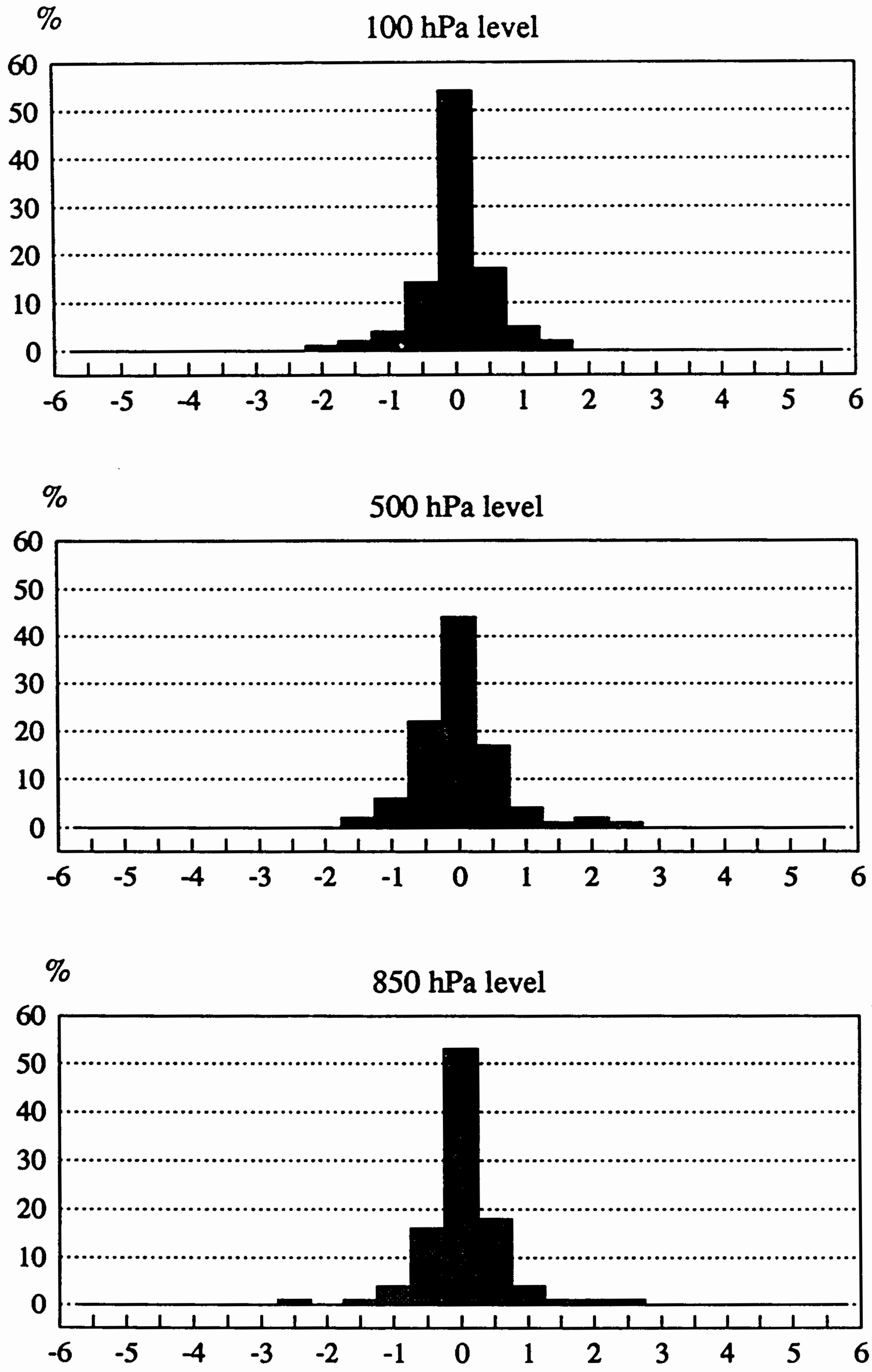


Fig. 39. Same as figure 37 except the meridional component of the wind.

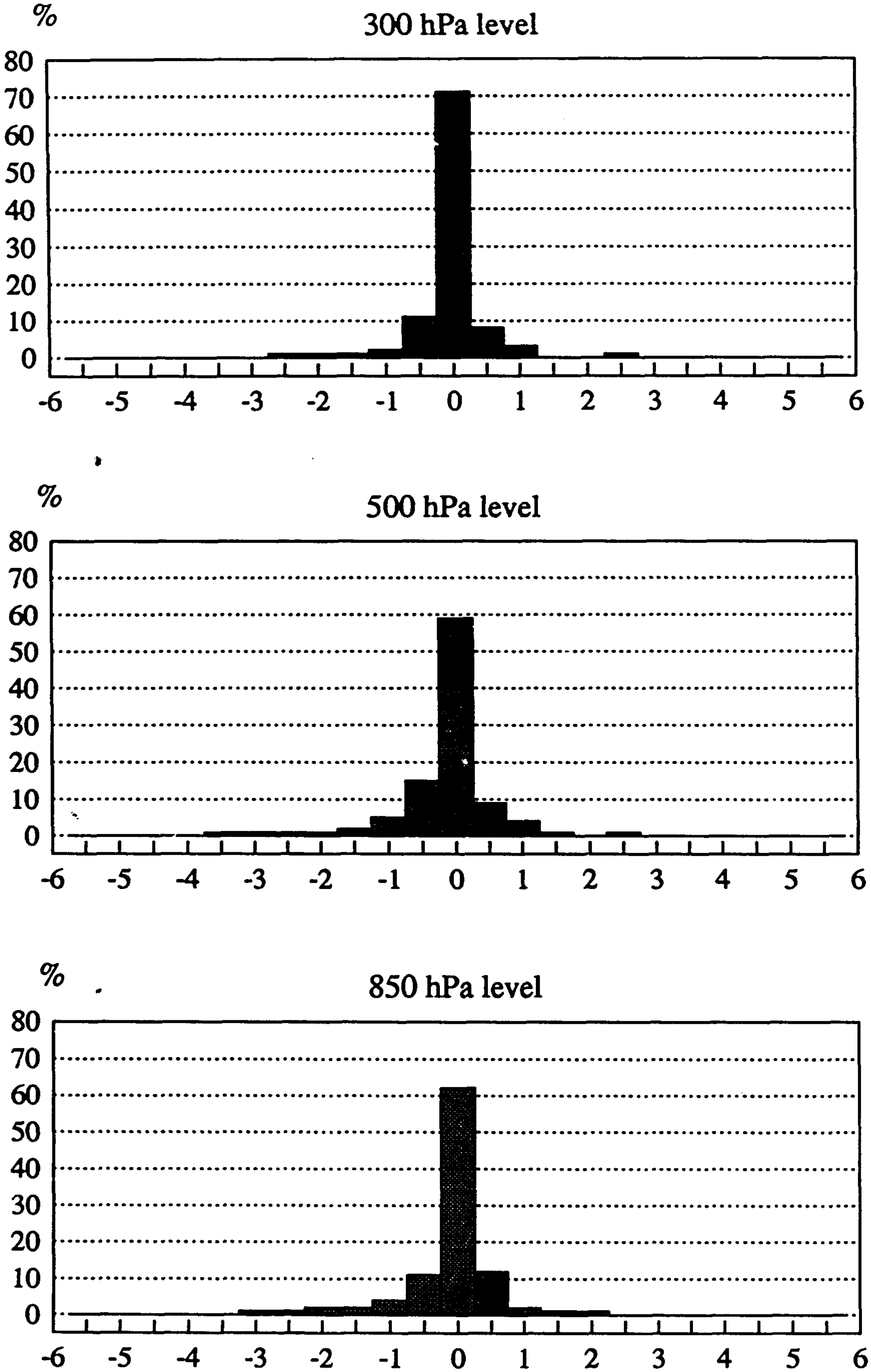


Fig. 40. Same as figure 37 except for dewpoint depression.

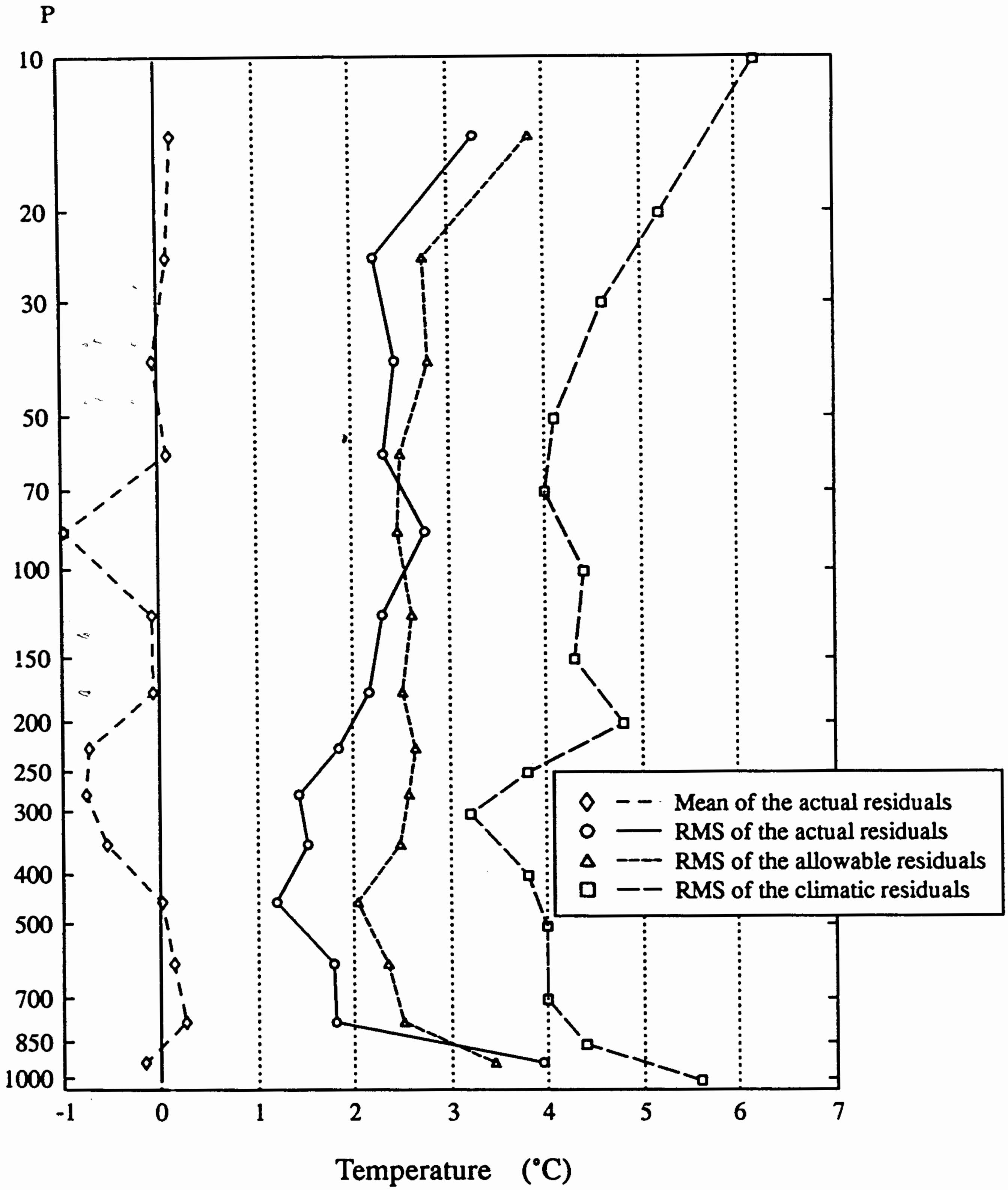


Fig. 41. Characteristics of vertical linear interpolation of temperature from mandatory levels to significant levels for a dataset of 759 stations from 00 UTC, 15 Jan 1989.

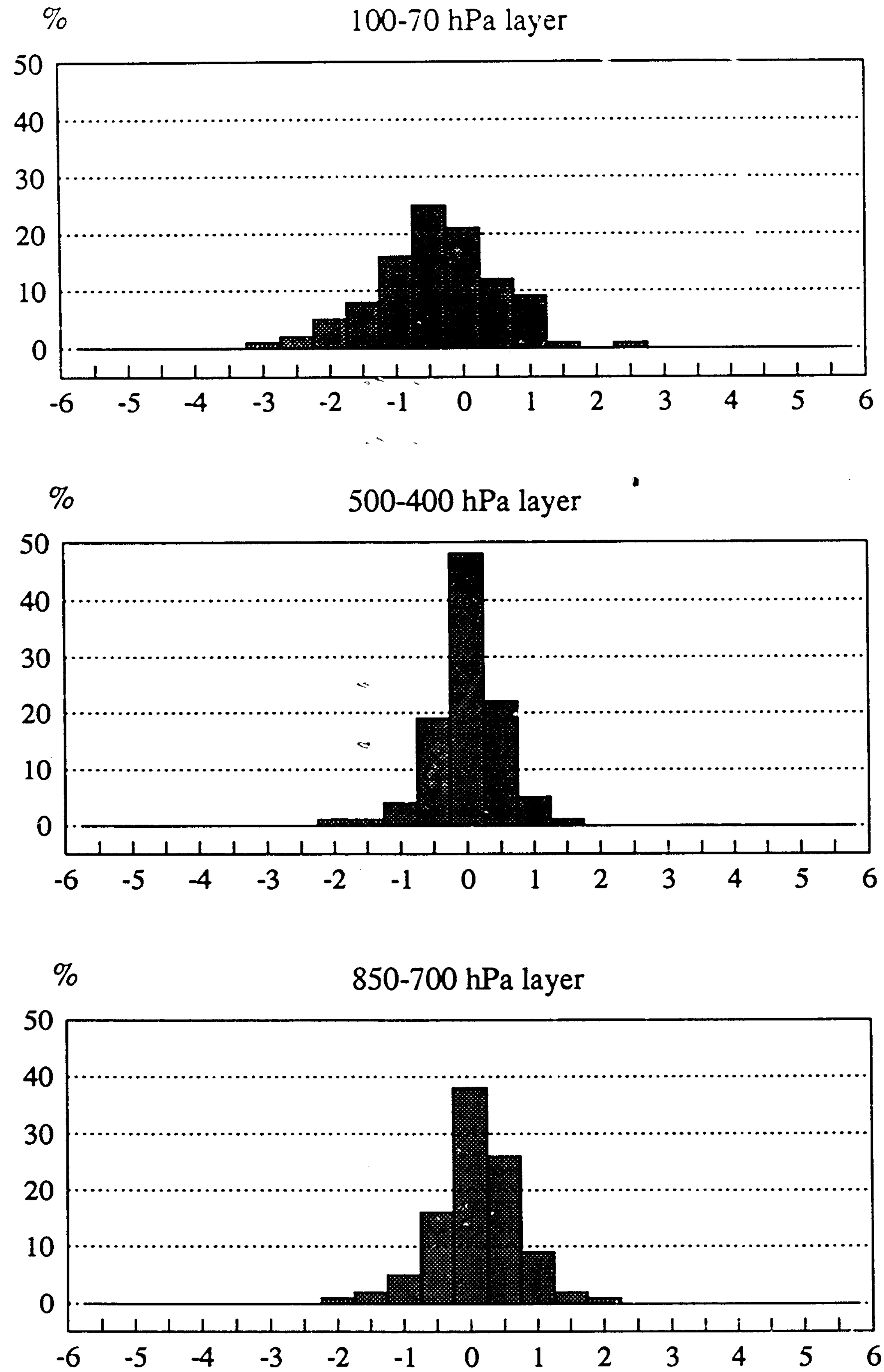


Fig. 42. Distribution of the normalized actual residuals from vertical linear interpolation of temperature from mandatory levels to significant levels for a dataset of 759 stations from 00 UTC, 15 Jan 1989.

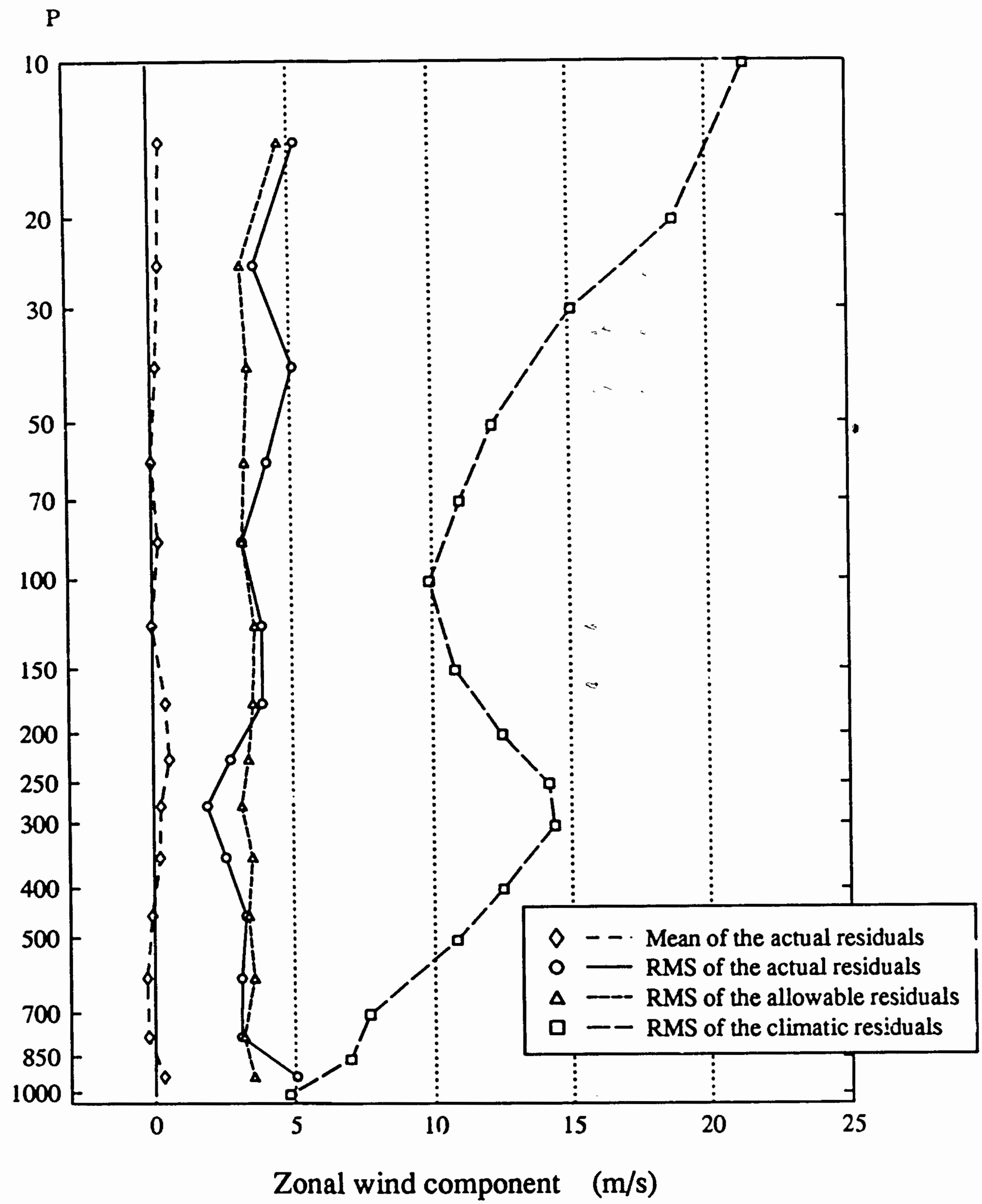


Fig. 43. Same as figure 41 except for the zonal wind component, U.

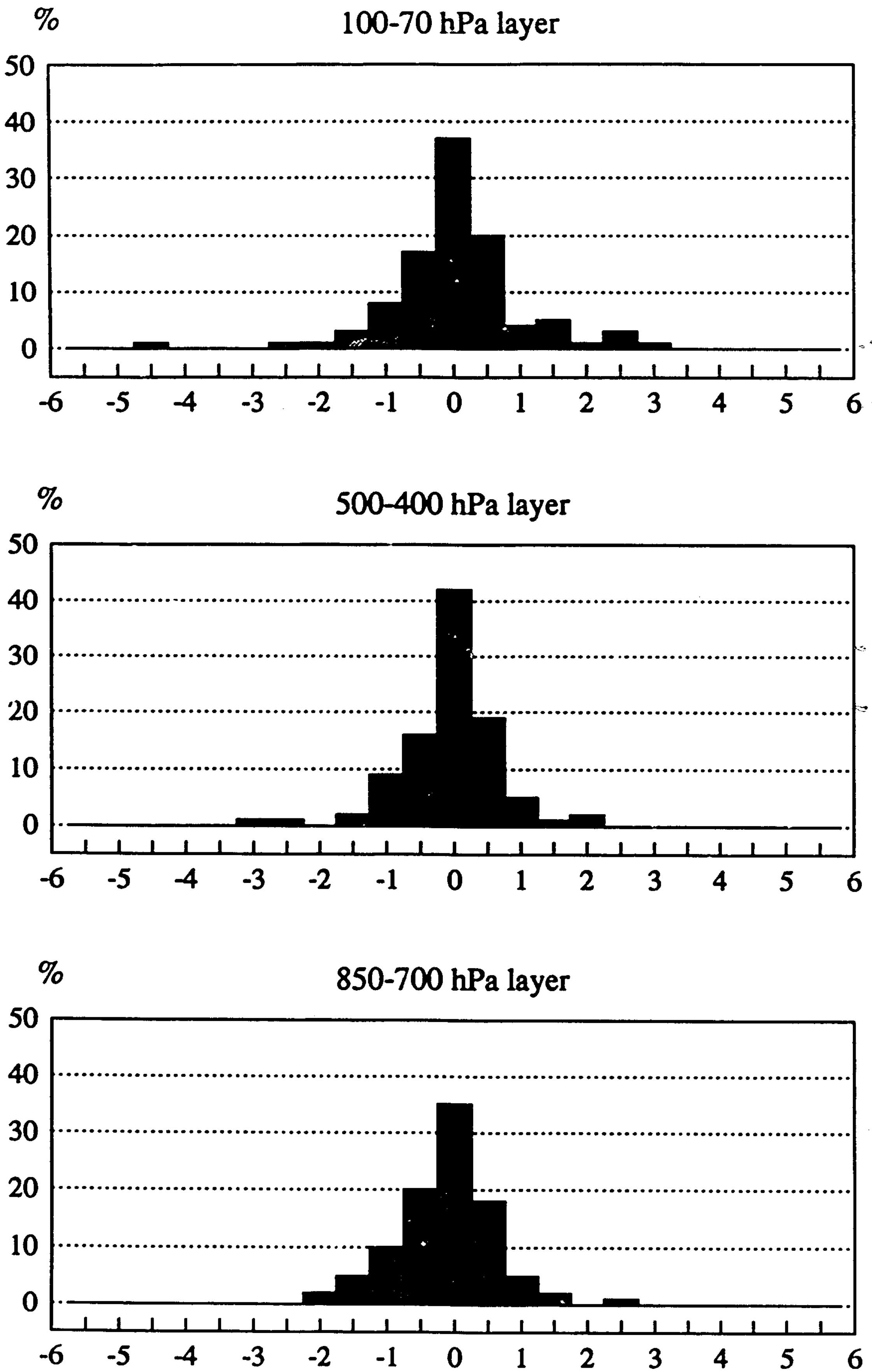


Fig. 44. Same as figure 42 except for the zonal wind component.

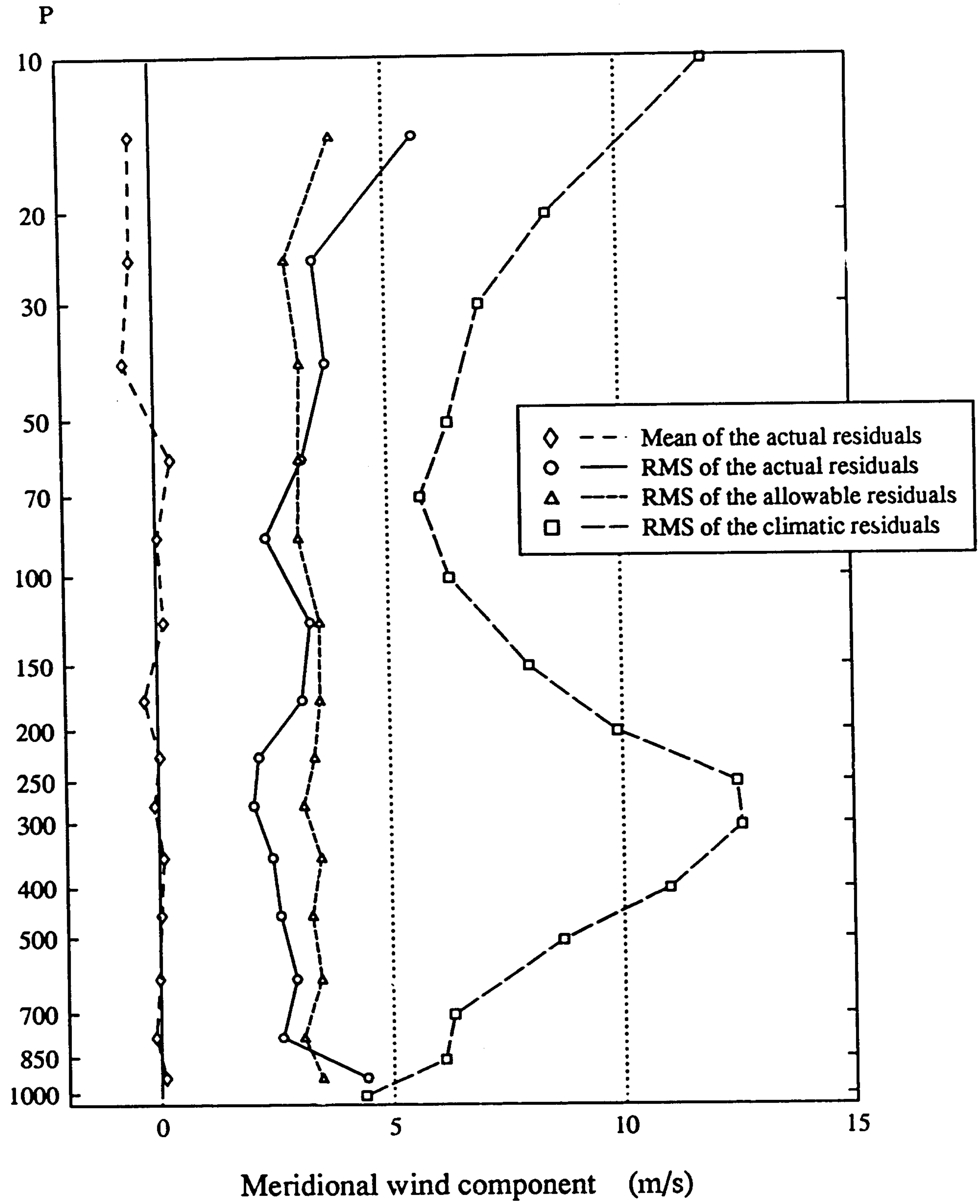


Fig. 45. Same as figure 41 except for the meridional component of the wind, V..

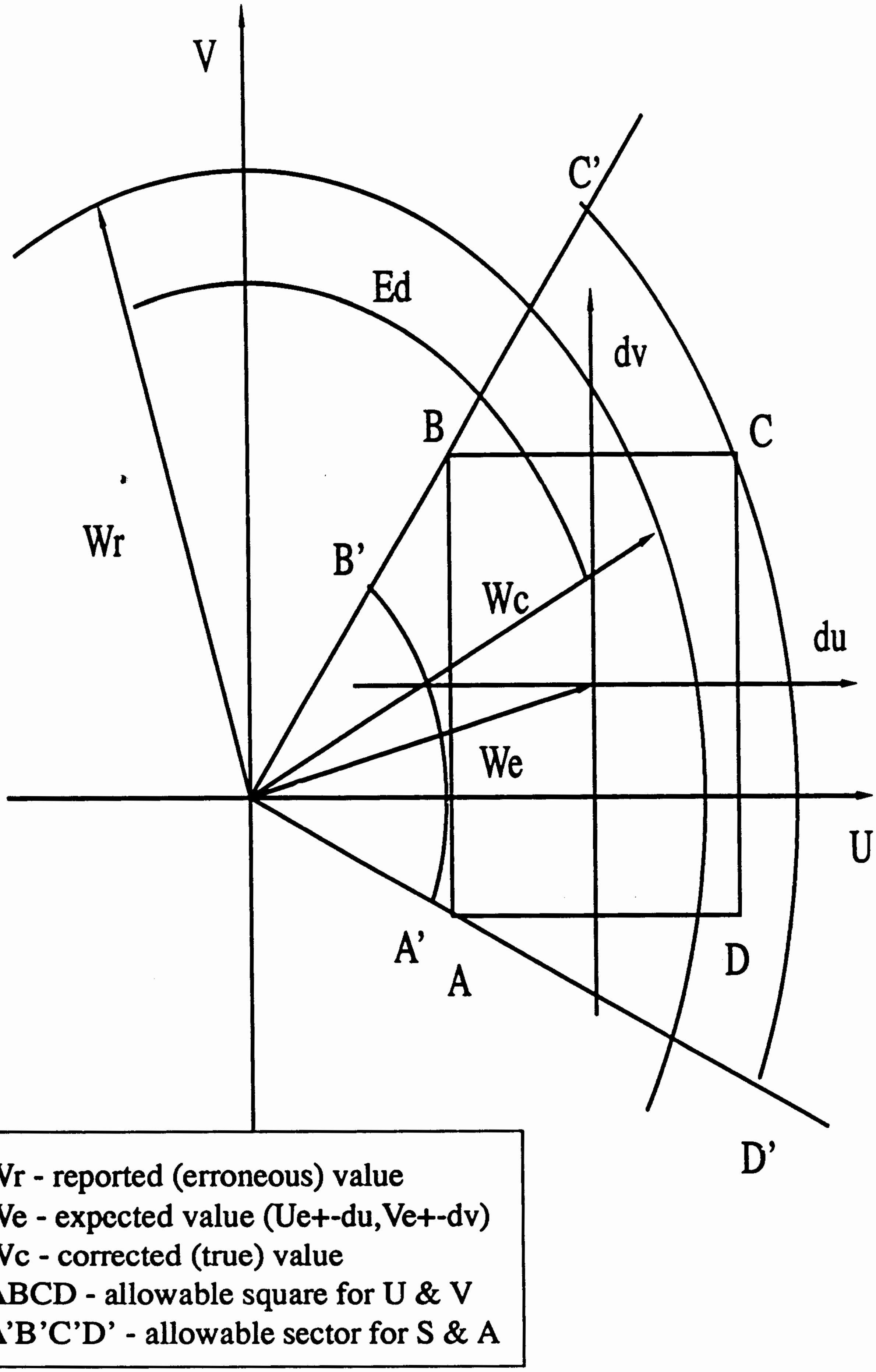
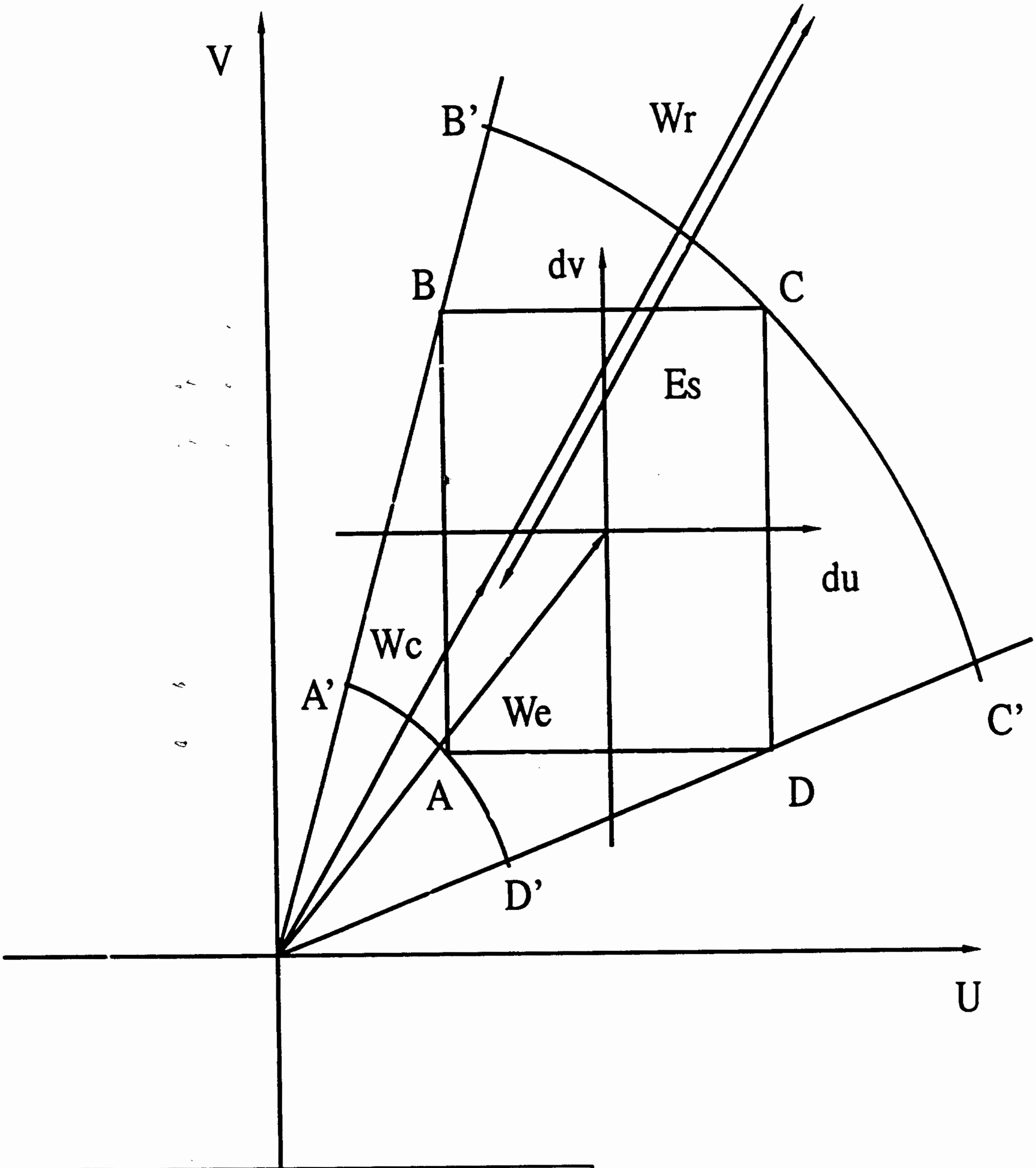


Fig. 49. Error (Ea) in wind direction



Wr - reported (erroneous) value
We - expected value (Ue+-du,Ve+-dv)
Wc - corrected (true) value
ABCD - allowable square for U & V
A'B'C'D' - allowable sector for S & A

Fig. 50. Error (Es) in wind speed